DOCUMENT RESUME

ED 123 063	SE 020 504
TITLE	A Report on Research into the Nuffield Advanced Chemistry Course and Examination, With Special Reference to Curriculum Evaluation and Development by Continuous Processes.
INSTITUTION	Lancaster Univ. (England).
PUB DATE	75
NOTE	191p.; Occasional marginal legibility; Prepared by Department of Educational Research
EDRS PRICE	MF-$0.83 HC-$10.03 Plus Postage
DESCRIPTORS	*Chemistry; Curriculum; *Curriculum Development; Curriculum Evaluation; *Educational Research; Science Course Improvement Project; Science Education; Secondary Education; *Secondary School Science; Tests
IDENTIFIERS	*Nuffield Project; Research Reports

ABSTRACT

This report summarizes an investigation to determine whether the outcome of examinations can be used as an aid to a continuous process of curriculum development, using the Nuffield A-level Chemistry Course and its system of assessment as a sample system. A description of the course and examination is followed by an explanation of how the data were collected and used to examine the relationship between question classification and student performance. Also included are analyses of the performance of various sub-groups of Nuffield chemistry candidates and teachers' and students' perceptions of some aspects of the course and examination. Curriculum evaluation and development are discussed as continuing processes. Also discussed are the kinds of evaluation that should prompt curriculum change. Appendices include questionnaires administered to teachers and students, sample examination questions, and notes on teacher assessment of student work. (MLH)

Documents acquired by ERIC include many informal unpublished materials not available from other sources. ERIC makes every effort to obtain the best copy available. Nevertheless, items of marginal reproducibility are often encountered and this affects the quality of the microfiche and hardcopy reproductions ERIC makes available via the ERIC Document Reproduction Service (EDRS). EDRS is not responsible for the quality of the original document. Reproductions supplied by EDRS are the best that can be made from the original.

ED123063

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

A REPORT ON RESEARCH INTO THE NUFFIELD ADVANCED CHEMISTRY COURSE AND EXAMINATION

WITH SPECIAL REFERENCE TO CURRICULUM EVALUATION AND DEVELOPMENT BY CONTINUOUS PROCESSES

BEST COPY AVAILABLE

SE020 504

A REPORT ON RESEARCH INTO THE

NUFFIELD ADVANCED CHEMISTRY COURSE AND EXAMINATION

With Special Reference to
Curriculum Evaluation and
Development by Continuous
Processes.

CONTENTS

PREFACE

INTRODUCTION

SECTION 1 The Nuffield Chemistry Course and Examination.

SECTION 2 Collection of Data.

SECTION 3 The Relationship between Question Classification and Student Performance.

SECTION 4 The Free Response Paper.

SECTION 5 The Assessment of Practical Work.

SECTION 6 The Performance of Various Sub-Groups of Nuffield Chemistry Candidates.

SECTION 7 Teachers' and Students' Perception of Some Aspects of the Course and Examination.

SECTION 8 Curriculum Evaluation and Development by Continuous Processes.

PREFACE

The detail of this report may be of interest to those teachers, examiners, and curriculum developers who are concerned with the Nuffield Advanced Chemistry course. The general findings, particularly those which deal with the principles of continuous curriculum evaluation and development, are intended for a wider readership. The latter may find that the Introduction and the final section, Curriculum Evaluation and Development by Continuous Processes, of most interest to them, using the body of the report for reference.

The work owes much to the assistance given by many organisations and individuals. Our thanks are extended to The Schools Council, The Nuffield Foundation, and the London University School Examinations Department for providing money and other resources; to The Department of Educational Research of the University of Lancaster for providing a home for the project: to Mr. Geoffrey Sleightholme and Mr. Leslie Stringer who contributed to two important parts of the research; to the teachers for giving an almost unanimous response to our questionnaires; and - above all - to the thousands of students whose efforts during the course and in their examinations provided most of the raw material for our research.

J.C. MATHEWS - Project Director
J.R. LEECE - Project Research Officer

LIST OF APPENDICES

APPENDIX

1. Examples of Types of Questions Used in the Nuffield 'A' level Chemistry Examinations.

2. Notes on the Assessment of Practical Work by Teachers (up to 1973).

3. An Example of Question Classification and Examination Specification.

4. Mark Sheets for Abstracting Scores.

5. Example of a Structured Question Designed for Pre-coding Marks.

6. Some Ancillary Data.

7. Questionnaires.

8. Measurement of Objectives of Practical Work in Conjunction with Particular Experiments.

INTRODUCTION

The great wave of curriculum development which started in the United Kingdom in the 1960's was accompanied by an equally strong move towards a reformation of evaluation and assessment techniques. Changes in evaluation and assessment manifested themselves in two main areas: the first in use of trials in schools of the curriculum materials during the formative period of the curriculum development, and the second in changes in techniques of assessment of student attainment at the end of the course. A comparison of present-day examinations in the General Certificate of Education with those of the 1950's reveals some striking changes which may be summarised as follows.

The use of new techniques, in particular the use of fixed response questions (objective tests), structured questions, multiple marking by impression, and the introduction of an element of internal assessment by teachers.

The introduction of examination specifications. These specifications are statements in a quantitative form of what the examination was intended to test, not only in terms of its subject content, but in terms of the objectives and other features of the curriculum.

An extension of teacher involvement. This showed itself partly in the element of internal assessment, partly in the increased use of teachers in the construction of the written examination, and partly in closer consultation with teachers about the structure and operation of the examination.

The changes in curricula and assessment systems have tended to be achieved by batch processes, revolutionary rather than evolutionary, a three year period of development and trials followed by the publication of curricular materials and an assessment system in a form which was unlikely to be changed for some years.

The students themselves have also been processed in batches. Their attainment, as shown in their examination scripts or internally assessed record cards, has been used to grade them for despatch to the next stage of their careers and the evidence of their attainment then allowed to gather dust.

The state of discontent with the curricula and examinations in the 1960's had reached such a pitch that a revolutionary type of change was probably inevitable. But now that this phase is passing a continuous process of change is probably more appropriate, and this requires a system of monitoring the functions and the outcomes of the curriculum to provide information on the basis of which changes of a less abrupt nature can be made.

The main purpose of the investigation now reported was to determine whether the outcome of examinations could be used as an aid to a continuous process of curriculum development. To do so, it was necessary to investigate an assessment system which had been designed to match a particular curriculum; and the Nuffield 'A' level Chemistry Course and its system of assessment appeared to provide a suitable medium for research of this type.

It is not anticipated that all curricula and examinations could be monitored by exactly the same process as that appropriate to Nuffield 'A' level chemistry. Nevertheless, it is hoped that all the principles and some of the processes could be applied to the science subjects at least, and it would be rash to assume that none of them is applicable to the humanities and arts.

SECTION 1.

THE NUFFIELD 'A' LEVEL CHEMISTRY COURSE AND EXAMINATION

The Nuffield Advanced Chemistry Course and its Examination were developed together during the period 1965 to 1968, so that the structure of the examination matched the structure of the course as closely as possible. The examination itself introduced techniques of assessment which were new to science examinations at this level. This combination of new techniques and a close relationship between course and examination gave grounds for assuming that Nuffield 'A' level Chemistry would provide suitable material for the research.

Details of the course and its assessment system are set out in the publications listed in references 1 to 5. Briefly, it is a two year course of integrated practical and theoretical work. It consists of nineteen topics, integrated by seven major themes. Emphasis is on student activity rather than on passive learning - "You can find out only by doing chemistry, not by reading about it".[4] In addition to the basic course, students are expected to undertake a special study, of which - at the time of writing - there are five in all.

The Examination

The Examination consists of five distinct parts:-

(The numbering of the Papers was changed in 1973. In this report the numbering which was in use during the period under investigation, 1969-1972, is used. Examples of each type of question are to be found in Appendix 1.).

Paper 1 is an objective test comprising five different types of questions arranged in five sections: multiple choice, classification sets, situation sets, multiple completion, and relationship analysis. There are fifty questions in Paper 1, all of which are compulsory, and its duration is 75 minutes.

Paper 2 consists of eight or nine structured questions. Each question is sub-divided, giving a total of 45 to 50 sub-questions, all of which are compulsory. The duration of the Paper is 2 hours.

Paper 3A is used to test attainment in the Special Study. It is not a main subject of investivation by this Project.

Paper 3B consists of approximately 9 questions, of which candidates are expected to answer 3. The Paper invites imaginative responses and it is in this part of the examination that candidates' ability to communicate freely is tested.

Assessment of Practical Work

This is internally assessed by the teachers. The instructions for the assessment (up to 1973), together with a specimen record card, is given in Appendix 2 .

Marking

Paper 1 is marked by machine scanning, each correct response being awarded 1 mark. There is no penalty for incorrect responses.

Paper 2 is marked by a team of assistant examiners, using a detailed mark scheme. The mark scheme is finalised at a meeting of examiners after marking a sample of scripts. Chief examiners re-mark a sample of scripts from each of the assistants. The marks are recorded alongside each sub-question, totalled at the end of each question, and listed and totalled on a fly leaf at the end of each script.

In Paper 3B each question is graded (5-4-3-2-1-0) by impression by two examiners, the mark awarded being the sum of the two grades. The first examiner records his grades on a fly-leaf; no mark or correction is made on the answers themselves. If grades differ by more than two, the answer is re-marked by a chief examiner.

The Examination Specification

One of the reasons why the Nuffield A level chemistry examination was chosen as the subject of this research was that each question in Papers 1 and 2 is described by four dimensions: topic, theme, objective and activity. Since all the questions are compulsory, it should be possible to determine whether there was any correlation between the various dimensions and student performance.

The topics are the teaching units listed in the published materials:-

1. Amount of substance
2. Periodicity
3. Molecular weight, atomic weight and the Avogadro number.
4. Atomic structure.
5. The halogens and oxidation numbers.
6. The s-block elements and the acid/base concept.
7. Energy changes and bonding.
8. Structure and Bonding.
9. Carbon Chemistry - Part 1.
10. Intermolecular forces.
11. Solvation.
12. Equilibria (I)
13. Carbon Chemistry - Part 2.
14. Reaction Rates.
15. Equilibria (II): Redox and acid/base systems.
16. Some d-block elements.
17. Equilibrium and free energy.
18. Carbon compounds with large molecules.
19. Some p-block elements.

The themes have an integrating function and provide a description of the content of chemistry which is independent of specific topics and material. They are:-

1. Materials behave and are composed in a specific manner.
2. There are patterns in the behaviour and composition of materials.
3. The behaviour and composition of materials can be explained by their structure.
4. The speed with which chemical changes take place - kinetics.
5. How far chemical changes are completed - equilibria.
6. Changes in materials are accompanied by energy changes - energetics.
7. Chemistry influences mankind and the environment - applications.

(Theme 7 is rarely the subject of a question in Papers 1 or 2).

The abilities (objectives) describe the particular attributes of the student tested by the question. They are classified in accordance with the first four levels of Bloom's Taxonomy of Educational Objectives (Cognitive Domain) :-

1. Knowledge.
2. Understanding.
3. Application.
4. Analysis

The classification by activity is intended to reflect the type of learning activity which is appropriate to each question:-

1. Using specific changes in materials.
2. Using practical techniques.
3. Using patterns in the behaviour of materials.
4. Using measurements and calculations.
5. Using concepts.

Question writers classify each question in Paper 1 and each sub-question in Paper 2. The classification of Paper 1 questions is revised by a small group of reviewers and then by a chief examiner; the classification of Paper 2 questions is revised by a chief examiner only. An example of a final classification and an overall examination specification is given in Appendix 3.

The assessment of practical work has a three dimensional structure:- objective, areas of work, type of work.

The objectives are:-

1. Skill in observation
2. Ability to interpret observations.
3. Ability to plan experiments.
4. Manipulative skills.
5. Attitudes to practical work.

The Areas of work are:-

1. Changes in substances and patterns in changes in substances. (This a deliberately broad category to include most of the course not covered by 2, 3 and 4 below. Most of the assessed experiments are likely to come in this category.)

2. Equilibria.
3. Kinetics.
4. Energetics.

The Types of work are:- 1. Quantitative work,
2. Preparative work, 3. Qualitative work.

For the purpose of this research only the objectives were taken into account.

REFERENCES

1. Nuffield Advanced Science (Chemistry). Examinations and Assessment. Penguin. 1972.

2. Nuffield Advanced Science (Chemistry). Teachers' Guide I. Penguin, 1970.

3. Nuffield Advanced Science (Chemistry) Teachers' Guide II. Penguin, 1970.

4. Nuffield Advanced Science (Chemistry). Students' Book I. Penguin, 1970.

5. Nuffield Advanced Science (Chemistry). Students' Book II. Penguin, 1970.

6. BLOOM, B.S. Taxonomy of Educational Objectives: Handbook I Longmans. 1956.

SECTION 2

THE COLLECTION AND PROCESSING OF DATA

In the monitoring of curriculum and assessment systems the ease and speed with which data can be collected, processed and disseminated are important factors. This section describes how data was collected and processed for the present investigation.

Data from Papers 1 and 2.

In some respects the data from Papers 1 and 2 are of prime importance to the investigation. Not only do they supply almost 100 of the 110 (approximately) separate scores recorded for each candidate in the examination, they are scores on questions attempted by all candidates, no choice between questions being allowed in the two Papers. The large number of compulsory questions allows coverage of all the main categories into which questions are divided; without such coverage analyses of the type which are described later would not be possible. Indeed, monitoring of the type described in Section 8 is inconceivable without a large, finely divided, and compulsory element in the examination. This is not to say that Papers of the type illustrated by Papers 1 and 2 of the Nuffield examination should be included solely to allow this type of monitoring; such action would be difficult to defend; but the point is made that if such elements have already been included on fundamental educational grounds, they allow a type of monitoring which other forms of assessment do not.

Sampling

In the 1970 and 1971 examinations the number of candidates was such (694 and 1084 respectively) that sampling was not necessary. For various reasons, for example the prior removal of some scripts for other purposes, it was not possible to record data from every script: a sample of 86% was achieved in 1970 and 97% in 1971. In 1972 the total number of candidates was 2174 and a sample of one third of centres was randomly selected, giving rise to a sample of 33.7% of candidates.

Data Set 1 (Paper 1)

This data set comprises the score on each question in Paper 1 for all the candidates in the sample, together with the total score on the Paper. The candidates' responses to each question are read by a scanning machine from the answer sheets and marked by a computer which records the scores on a magnetic tape. Correct answers are scored 1, incorrect answers 0, and the total score is the number of correct responses.

This data set was available within hours of the receipt of all the answer sheets by the examining board. The rate controlling step is that between the examination and the receipt of answer sheets from all the schools.

Data Set 2 (Paper 2)

This data set comprises the score on each sub-question in

Paper 2, the total score on each question, and the total score for Paper 2.

The scores were recorded manually on the scripts in red ink by the examiners. The sub-question score in the margin, the total score for each full question at the end of the question and again on a fly-leaf, and the total score for the Paper on the fly-leaf and on the front of the script. It is apparent, therefore, that the abstraction of scores and part-scores involved manual transference from the scripts, to mark sheets (Appendix 4.) and thence to punched computer data cards.

The manual transference of scores to mark sheets is a big task. The process is typified by that for the 1972 examination:

A random selection of one-third of centres was done by the Projects Research Officer, but could be undertaken by clerical assistants, provided that clear instructions were given and a routine established.

Transference of scores to mark sheets was done by clerks after about one hour's instruction from the Research Officer, followed by supervision for the first day's work by the Research Officer who dealt with operational queries as they arose. The transference of data from 733 scripts to mark sheets occupied the time of one clerk for 88 hours, working at the rate of 7 hours each day. The time span for the operation was approximately two weeks. The next step was the transference of the data from the mark sheets to punched computer data cards by professional punched card operators. In 1970 and 1971 this punching was not subject to a punched card verification process. The subsequent computerised checking revealed an error rate of 60 (mixed coding and punching) out of 596 cards in 1970, and 107 coding errors and 39 punching errors out of 1046 cards in 1971. In 1972 the process was subjected to punched card verification and the use of a special coding form. This reduced punching errors to zero and coding errors to 18 out of 733 cards.

The time taken for both the initial punching and verification punching was approximately 9 hours for 733 mark sheets. It should be noted, however, that punched card verification checks the mark sheet to punched card operation; it does not check the transference of marks from script to mark sheet. For the latter, an error detection element must be built into the computer programme used for processing the data. There follows a visual check of the errors located by the computer programme against the original scripts. This factor must be variable, depending on the accuracy of the clerical workers.

In 1972 39 scripts had to be re-checked, the cards re-punched, and the computer programme re-run to check again. The manual part of this operation took approximately 3 to 4 hours. Of course, scrupulous checking of the coding part of the process may not be necessary, particularly when dealing with large numbers of scripts when a certain level of error could be tolerated without significantly altering the results.

The abstraction of Data Set 2 is a much lengthier process than that for Data Set 1 and, clearly, it could not take place in its present form at the height of the Summer examination season. It had to wait until the Autumn and this meant that the earliest that the data could be processed and the findings disseminated was some six months after the examination had been taken and three months after the examining board's Advisory Panel had met to discuss the outcome of the examination and the examiners' reports. For long-term, periodic monitoring of the examination and curriculum so long an interval is, perhaps, not important, but swifter collection and processing of data is possible (see p. 135) and could be available to advise examiners before their awarding and standardising meetings. Furthermore, if, as will be suggested later, an information service about examination outcomes direct to schools is thought to be desirable, speed of reporting would be an advantage.

Data Set 3 (Paper 3B)

This data set comprises the two scores awarded for each question attempted in Paper 3B and the total score for the Paper.

These scores were recorded by each examiner on the fly-leaf of the scripts, but by special arrangement with the examining board the marks were also recorded on a mark sheet (Appendix 4.) designed specially for this research. Computer data cards were punched from the mark sheets and then processed. Thus, no manual transference of marks from scripts was required and the findings (see Section 4.) were available to the chief examiners in time for their award and standardising meeting; an event which would not have been possible if the scores had had to be abstracted from the scripts by clerical assistants.

Data Set 4. This data set comprises ancillary data as well as examination data and was recorded on a mark sheet.

1. The score awarded to each candidate by the teacher for each of the five objectives of the practical work, together with the total score for the practical work. These scores were transferred by clerical assistants from the candidates' record cards for practical work (Appendix 2).

2. The score on the set of questions in Papers 1 and 2, which was used to moderate the teachers' practical mark, was recorded on the same mark sheet as that used in 1. The source of this data was the mark sheets used by the examiners to abstract the sub-set of practical work scores while marking Papers 1 and 2.

3. The teacher's estimate of each candidate's 'A' level grade (not available in 1972) was recorded from the returns made to the examining board from each school.

4. The teaching set in which the candidate worked for practical work was transferred to the mark sheet from the record cards.

5. The candidate's sex, and the grades obtained in the other 'A' level subjects taken at the same time as Nuffield Chemistry were recorded on the mark sheet, the sources of data being the examination entry form and the published list of grade awards supplied by each Examining Board to the Undergraduate Admission Officer of Lancaster University.

The data recorded in Set 4 is heterogeneous in nature, coming from three separate sources. Guided clerical assistance is required and it is unlikely that the data could be processed in a shorter time than 12 weeks after the examination, this being determined by the time taken to publish the results. The estimated time required to abstract data in this Set when it has been assembled in its raw form for 733 candidates is 100 man hours of clerical work.

Data Set 5

This comprises a list of all candidates with the total score which each gained in Paper 1, 2, 3A, 3B, practical work (moderated score), and the examination grand total, that is the foregoing five scores combined. This data was available on magnetic tape recorded as a routine process by the examination board.

Combining the Data Sets

After verification, the four data sets were combined by identifying each candidate's separate data set, and then joining them by a computer programme, the output of which was recorded on a single magnetic tape.

Data from Questionnaires

As part of the research two questionnaires were administered. Each was in two parts, one directed to students and the other to teachers. The first questionnaire related to the internal assessment of practical work and is considered in Section 5. The second related to some aspects of the teachers' and students' perception of the Nuffield course and examination and is considered in Section 7. The responses were pre-coded on the questionnaires so that the data could be transferred directly to punched cards for subsequent processing.

Data Processing

It was realised at the outset that computerised techniques were essential in view of the large amount of information which had to be processed, and during the research many computer programmes were devised to put the information into a concise and meaningful form. The type of transformation applied to the data is outlined in each of the appropriate sections of the report. The details of the programming are too complex to be reported here. The main point to be made at this stage is that effective, continuous curriculum evaluation is only possible if the maximum use is made of computerized techniques.

The writing of appropriate programmes is very time consuming in the first instance; thereafter the main rate determining stage is not the processing of information, but its collection. There is no doubt that changes in the collection of information will be necessary if this form of monitoring is to become a routine procedure; and some suggestions for such changes are made in the final section of this report.

SECTION 3.

THE RELATIONSHIP BETWEEN QUESTION CLASSIFICATION AND STUDENT PERFORMANCE

In essence this part of the project was designed to find out whether any of the characteristics of the course or the examination were giving rise to relatively high or low performance on the part of the students. If one or more of these characteristics were found to be related to abnormal student performance, the attention of those responsible for the design of the course and the examination could be drawn to them and action taken, if action were thought to be necessary. In other words, the outcome of the research could serve as a basis for decision making.

In the tables which follow, the mean value of question facility indices (F) and the mean value of question discrimination indices (D) for each characteristic of the course or examination are listed, together with the number of questions in each category (N). For example, the first table compares the facility and discrimination of Paper 1 questions with the facility and discrimination of Paper 2 questions. In this section only those relationships which are directly relevant to the main objectives of the research are given; the computer print-outs contain many more which are not so relevant and a selection is given in Appendix 6 .

F = Mean score divided by maximum score on each question or sub-question.

D = Product moment correlation of each question or sub-question with the examination grand total. (The correlations with Paper totals and teacher estimates were also calculated but not used).

COMPARISON OF PERFORMANCE IN PAPERS 1 and 2

(The Objective and Structured Tests)

TABLE 3.1

MEAN VALUES OF QUESTION FACILITY (F) AND DISCRIMINATION (D) INDICES

PAPER		1970	1971	1972	YEARS COMBINED
1	F	0.50	0.51	0.55	0.52
	D	0.26	0.24	0.30	0.27
	(N)	(50)	(50)	(50)	(150)
2	F	0.51	0.57	0.51	0.53
	D	0.35	0.35	0.35	0.35
	(N)	(44)	(45)	(46)	(135)

PAPERS 1 and 2 ALL QUESTIONS	F	0.50	0.54	0.53	0.53
	D	0.30	0.30	0.33	0.31
	(N)	(94)	(95)	(96)	(285)

The most significant conclusion arising from Table 3.1 is that Paper 2 items discriminate between candidates better than Paper 1 items do. This is so for all three bases for measuring discriminations: Paper 1 or 2, the examination as a whole, and the teachers' estimates. In each year the mean value of (D) for Paper 1 is lower than that for Paper 2: (The F-ratio for one way analysis of variance showed that this difference is significant at the 1% level for the combined years and for years 1970 and 1971: in 1972 it is at the 5% level).

In contrast to the marked difference in discriminating power between the two Papers, the mean values for the facilities of the questions in Papers 1 and 2 are not significantly different from each other. These findings could well be noted by those who seek to extend the ability range which can be tested by a single examination.

COMPARISON OF PERFORMANCE IN THE VARIOUS TYPES OF QUESTIONS IN THE OBJECTIVE TEST (See Appendix 1.)

TABLE 3.2

QUESTION TYPE		1970	1971	1972	YEARS COMBINED
MULTIPLE CHOICE	F	0.52	0.53	0.55	0.53
	D	0.27	0.21	0.33	0.27
	(N)	(19)	(15)	(17)	(51)
MULTIPLE COMPLETION	F	0.45	0.42	0.55	0.47
	D	0.21	0.22	0.33	0.25
	(N)	(15)	(16)	(12)	(43)
CLASSIFICATION SET	F	0.55	0.75	0.63	0.65
	D	0.36	0.31	0.32	0.36
	(N)	(5)	(7)	(7)	(19)
SITUATION SET	F	0.49	0.45	0.56	0.51
	D	0.25	0.30	0.21	0.25
	(N)	(5)	(6)	(8)	(19)
ASSERTION REASON	F	0.53	0.49	0.47	0.50
	D	0.32	0.27	0.27	0.29
	(N)	(6)	(6)	(6)	(18)

PAPER 1.	F	0.50	0.51	0.55	0.52
	D	0.26	0.24	0.30	0.27
	(N)	(50)	(50)	(50)	(150)

The following inferences arise from this analysis:-

1. The questions displaying the highest facility are the classification sets and those of the lowest facility are the multiple completion type. The difference between the two is significant at the 1% level. (Scheffé S - Test).

2. The questions in the classification sets show higher discrimination indices than those of all the other questions in Paper 1. The difference of the mean values of D between classification sets and Paper 1 as a whole is significant at the 5% level (T - test for all Paper 1 questions and therefore slightly biased by the inclusion of classification questions in the mean value of D).

These two findings may have some interest for test constructors. At first sight the high facility values of the classification questions might be thought to arise from the frequent occurrence of non-functioning responses in this type of question and, consequently, a higher guessing factor; but the high discrimination values provide contrary evidence. Classification questions are comparatively easy to write; they are economical of time and space, and they could be regarded as strong contributors to the validity of chemistry examinations since classification is a major activity in chemistry and education in chemistry. At the same time, the questions give rise to a comparatively high proportion of correct responses and, therefore, are likely to be regarded as easy by teachers and candidates.

25. COMPARISON OF PERFORMANCE IN THE TOPICS OF THE COURSE

TABLE 3.3

TOPIC		1970	1971	1972	YEARS COMBINED
2	F	0.45	0.73	0.47	0.52
	D	0.17	0.28	0.39	0.24
	(N)	(5)	(2)	(2)	(9)
3	F	0.50	0.54	0.45	0.50
	D	0.28	0.37	0.33	0.33
	(N)	(3)	(5)	(5)	(13)
4	F	0.44	0.67	0.59	0.57
	D	0.32	0.28	0.25	0.28
	(N)	(4)	(5)	(3)	(12)
5	F	0.64	0.70	0.68	0.68
	D	0.24	0.32	0.28	0.29
	(N)	(3)	(5)	(7)	(15)
6	F	0.60	0.37	0.49	0.49
	D	0.10	0.29	0.20	0.23
	(N)	(4)	(4)	(2)	(10)
7	F	0.49	0.69	0.60	0.59
	D	0.33	0.34	0.31	0.32
	(N)	(5)	(5)	(11)	(21)
8	F	0.56	0.52	0.59	0.55
	D	0.34	0.26	0.27	0.29
	(N)	(5)	(7.	(5)	(17)

TABLE 3.3 (Contd.)

TOPIC		1970	1971	1972	YEARS COMBINED
9	F	0.47	0.47	0.44	0.46
	D	0.25	0.31	0.23	0.27
	(N)	(14)	(14)	(10)	(38)
10	F	0.53	0.62	0.09	0.55
	D	0.33	0.15	0.33	0.22
	(N)	(4)	(8)	(1)	(13)
11	F	0.48	0.31	0.30	0.43
	D	0.45	0.14	0.12	0.34
	(N)	(6)	(2)	(1)	(9)
12	F	0.42	0.49	0.55	0.50
	D	0.37	0.27	0.34	0.33
	(N)	(4)	(3)	(6)	(13)
13	F	0.58	0.44	0.52	0.52
	D	0.21	0.37	0.37	0.32
	(N)	(4)	(3)	(5)	(12)
14	F	0.58	0.53	0.71	0.58
	D	0.32	0.33	0.40	0.34
	(N)	(6)	(9)	(3)	(18)
15	F	0.50	0.55	0.33	0.46
	D	0.35	0.37	0.36	0.36
	(N)	(4)	(7)	(6)	(17)
16	F	0.32	0.51	0.49	0.44
	D	0.39	0.21	0.34	0.34
	(N)	(6)	(3)	(9)	(18)

TABLE 3.3 (Contd.)

TOPIC		1970	1971	1972	YEARS COMBINED
17	F	0.47	0.50	0.60	0.50
	D	0.32	0.24	0.34	0.29
	(N)	(4)	(4)	(1)	(9)
18	F	0.58	0.41	0.70	0.55
	D	0.30	0.36	0.27	0.31
	(N)	(2)	(6)	(5)	(13)
19	F	0.67	0.73	0.55	0.63
	D	0.27	0.33	0.40	0.35
	(N)	(3)	(3)	(5)	(11)

PAPERS 1 and 2 ALL QUESTIONS	F	0.50	0.54	0.53	0.53
	D	0.30	0.30	0.33	0.31
	(N)	(94)	(95)	(96)	(285)

It has to be noted that the classification by Topics of the curriculum and the examination questions is a rather fine division of the material. Although in each year all the Topics (except Topic 1) are represented in Papers 1 and 2 combined, giving an indication of the content validity of the Papers, the actual number of questions in a particular Topic may be quite small - Topic 2 for example.

This makes it unlikely that educationally meaningful differences which are both statistically significant and meaningful in the facility and discrimination of questions between Topics will be found. Sometimes such a difference can be found; for example the discrimination indices of Topic 9 questions in 1972 are significantly less than those of the Paper 1 and 2 questions as a whole in that year (T - test 5% significance). But in general it seems that, with tests of comparable numbers of items, a combination of three years of results is required before trends in facilities and discriminations can be shown to be statistically significant.

From the values of F and D in the combined years it seems that the only significant differences between Topics are as follows:-

Facility (F)

1. Topic 5 questions (F = 0.68) are the easiest, the difference between the mean facility of these questions and that of all the Paper 1 and 2 questions (0.53) is significant at the 1% level (T - test).

   Questions in some other Topics, for example 4, 7 and 14, tend to be easy, but the degree and significance of the differences are such that they do not give cause for concern or serve as a basis for action.

2. Only Topics 9 and 16 are significantly more difficult than the combined Papers at the 10% level (T - test). The lowest mean facility index is actually found in Topic 11

questions; but, because of the small number of questions (9) the difference has less statistical significance. It is also worth noting that the combined mean facility of Topic 16 questions is largely caused by one particularly difficult year (1970)

Topic 9 questions (F = 0.46), however, are worthy of further consideration. Topic 9 is by far the most heavily weighted Topic; it comprises a clearly defined area of chemistry (organic chemistry) and the low value of the mean facility is almost constant from year to year; furthermore the mean discrimination shows the same trend (see below).

So the main finding about the relative facility of questions classified by Topic, is that Topic 5 questions are considerably easier than Topic 9 questions to an extent which is sufficient to warrant serious consideration and possible action. The implications of this are discussed later in this Section.

Discrimination (D)

The questions which show the highest discrimination, based on the examination as a whole, are those of Topic 15 (D= 0.36) (T - test 10% level). There are three Topics which show a significantly low discrimination (T - test 5% level), 6, 9 and 10 (D = 0.23, 0.27, and 0.22, respectively).

Implications of the Findings from Table 3.3.

The first thing to be said is that most questions function in a very similar way from Topic to Topic and there is little cause for concern or action. The exception lies in Topics 5 and 9, and the question must be asked why do they differ substantially in the mean value of the question facility indices? It is at this point that one must leave the statistical analysis and seek other criteria as a basis for judgement. Analysis reveals areas of underperformance (Topic 9) and overperformance (Topic 5) but it does not reveal causes. The cause of differing performance could lie in one or more of the following:-

1. The subject matter of the Topic is intrinsically easier or harder for students to learn.
2. The subject matter is intrinsically easier or harder for teachers to teach.
3. The examination questions themselves are requiring too low or too high a standard of student ability.

It is suggested that these could be the main causes. There may be others: if Topic 9 questions, for example, were all of the multiple completion type then this might be the reason; students and teachers may find a Topic boring rather than difficult; a Topic at the end of the course may be hurriedly treated, and so on.

Topic 5 is a study of the halogens, largely experimental, using the study to introduce the concept of oxidation number. It seems, if one uses examination outcomes as the criterion, that both teachers and students find the material of the Topic and the concept of oxidation number easy relative to other parts of the course. What, if anything, should be done about it?

One course of action, in a revision of the curriculum, would be to increase the demands made on teachers and students by extending the study of the halogens and of the oxidation number concept by increasing the amount of material or by extending the conceptual content so that greater intellectual ability is required for complete understanding.

Another course of action, would be to write more difficult examination questions on this topic, possibly by using multiple completion questions in Paper 1. Both these courses of action could be expected to lower the standard of performance of the students so that it was nearer to that in the other parts of the examination. But then one must ask, why should students be subjected to a more extensive curriculum or more difficult questions just because they do better than those who constructed the curriculum and examinations.expected? It might be more appropriate, then, to make no changes and simply be thankful that at least one aspect of chemistry is being well taught and learned. However, with a Topic, such as Topic 9, which is giving rise to consistent underperformance, the findings may call for different action. Topic 9 provides an introduction to the chemistry of carbon compounds; it contains a good deal of application of structural and reaction mechanism concepts, but, in general, it is concerned with the behaviour and nomenclature of carbon compounds and is not far removed from the traditional study of Organic Chemistry. In Further and Higher Education organic chemistry has normally been an equal partner in the tripartite division, organic, physical and inorganic chemistry. The importance of this division may be disputed; nevertheless, the discovery of relatively low performance by students in the main Topic on carbon chemistry must prompt a search for reasons and point to a possible need for changes.

Identification of the problem is only the first step. That there is a problem with Topic 9 there can be little doubt, but where does the solution lie? Is it with the teachers, or the students, or the curriculum designers, or the examiners, or some combination of those groups? The present research can do no more than suggest lines for further action: perhaps a questionnaire to students and teachers, or a more detailed analysis of Topic 9 questions, or a review of the published materials for this Topic.

It is interesting to note that the other two Topics which deal with carbon chemistry, 13 and 18, do not show underperformance by the candidates; nor does the area of carbon chemistry as a whole (see below). Conjecture might be dangerous at this stage but could the answer lie in the fact that both topics 13 and 18 have a considerable orientation towards the applications of organic chemistry in the practical work?

With regard to the different mean discrimination indices, those of Topic 15 (Equilibria: redox and acid-base systems) are the only ones which call for comment. It is significant that Topic 15 is strongly physical and mathematical. It is, perhaps, too early to speculate as to why questions on this Topic form a group of questions which behave most like these of the examination as a whole, but reference will be made elsewhere (Section 6.   ) to the links between Nuffield Chemistry and the supporting disciplines. In passing, it should also be noted that Topic 9 which is non-mathematical and is concerned with organic rather than physical chemistry has a mean value of discrimination indices significantly lower than that of all questions in Papers 1 and 2 (T - test 5%). A straight comparison of discrimination indices between Topics 9 and 15 shows that the former are significantly lower at the 1% level (T - test).

COMPARISON OF PERFORMANCE IN GROUPED TOPICS

On the supposition that the large number of Topics (19) might prove to be too fine a division of the subject matter of the examination for meaningful analysis, the Topics were grouped into Areas (see below   ), but the subsequent analysis (Table 3.4) shows that, although there is some fluctuation from year to year, there is little significant difference between the F and D values of questions grouped in this way.

AREA

1. Organic - Topics 9, 13, 18.
2. Intermolecular Forces - Topics 10, 11.
3. Equilibria - Topics 12, 15, 17.
4. Inorganic - Topics 2, 5, 6, 16, 19.
5. Amount of Substance - Topics 1 and 3.
6. Structure and Bonding - Topics, 4, 7, 8.
7. Kinetics - Topic 14.

TABLE 3.4

AREA		1970	1971	1972	YEARS COMBINED
1	F	0.51	0.45	0.51	0.49
	R	0.25	0.33	0.33	0.31
	(N)	(21)	(23)	(27)	(71)
2	F	0.50	0.56	0.20	0.50
	R	0.41	0.15	0.23	0.27
	(N)	(10)	(10)	(2)	(22)
3	F	0.46	0.52	0.46	0.48
	R	0.34	0.31	0.35	0.34
	(N)	(12)	(14)	(13)	(39)
4	F	0.50	0.60	0.55	0.55
	R	0.27	0.29	0.33	0.30
	(N)	(23)	(17)	(26)	(66)
5	F	0.50	0.54	0.49	0.51
	R	0.28	0.37	0.35	0.34
	(N)	(3)	(5)	(6)	(14)
6	F	0.50	0.61	0.59	0.57
	R	0.33	0.29	0.29	0.30
	(N)	(14)	(17)	(19)	(50)
7	F	0.58	0.53	0.71	0.58
	R	0.32	0.33	0.40	0.34
	(N)	(6)	(9)	(3)	(18)
PAPERS 1 and 2 ALL QUESTIONS	F	0.50	0.54	0.53	0.53
	D	0.30	0.30	0.33	0.31
	(N)	(94)	(95)	(96)	(285)

COMPARISON OF PERFORMANCE IN THEMES

TABLE 3.5

THEMES		1970	1971	1972	YEARS COMBINED
1	F	0.49	0.53	0.46	0.49
	D	0.26	0.23	0.30	0.27
	(N)	(30)	(25)	(28)	(83)
2	F	0.53	0.60	0.59	0.57
	D	0.28	0.33	0.40	0.34
	(N)	(17)	(20)	(21)	(58)
3	F	0.50	0.50	0.52	0.51
	D	0.36	0.30	0.30	0.32
	(N)	(19)	(23)	(20)	(62)
4	F	0.62	0.53	0.71	0.59
	D	0.31	0.33	0.40	0.33
	(N)	(7)	(9)	(3)	(19)
5	F	0,46	0.55	0.45	0.48
	D	0.36	0.37	0.34	0.35
	(N)	(12)	(10)	(11)	(33)
6	F	0.46	0.54	0.60	0.54
	D	0.34	0.26	0.30	0.30
	(N)	(8)	(8)	(12)	(28)
PAPERS 1 and 2 ALL QUESTIONS	F	0.50	0.54	0.53	0.53
	D	0.30	0.30	0.33	0.31
	(N)	(94)	(95)	(96)	(285)

The Table shows that questions on Themes 2 and 4 give rise to the highest performance by candidates (F = 0.57 and 0.59), Themes 1 and 5 the lowest (F - 0.49 and 0.48), the difference between the high and low performance is statistically significant at the 5% level (T - test). In other words, candidates find the questions which deal with patterns in behaviour of materials easier than those which deal with specific behaviour of materials. And questions dealing with how fast reactions go prove easier than those dealing with how far reactions go.

With regard to the discrimination indices, the mean value of D for Theme 1 questions is significantly lower (D - 0.27) than that of all the questions in Papers 1 and 2 (T - test, 1% level), and the mean value of D for Theme 5 questions (D = 0.35) is significantly higher (T - test, 5% level).

These findings are similar to those which distinguish between Topics 9 and 15 (pp. 32-33); but they are at a more general level, because Theme 1 includes the specific behaviour of both organic and inorganic substances and Theme 5 includes equilibria concepts in Topics 12 and 17 as well as Topic 15. It seems, therefore, that although questions on the two Themes are almost equally difficult, those on Theme 5 discriminate between candidates in a manner which is closer to that of the examination as a whole. In still more general terms it could be deduced that the examination as a whole discriminates in a way which indicates that it has a bias towards physical and mathematical concepts and away from specific behaviour of materials.

COMPARISON OF PERFORMANCE IN ACTIVITIES

TABLE 3.6

ACTIVITY		1970	1971	1972	YEARS COMBINED
1	F	0.48	0.50	0.59	0.53
	D	0.32	0.28	0.35	0.32
	(N)	(16)	(11)	(18)	(45)
2	F	0.49	0.54	0.48	0.50
	D	0.26	0.19	0.27	0.22
	(N)	(20)	(20)	(20)	(60)
3	F	0.55	0.55	0.55	0.55
	D	0.27	0.34	0.32	0.31
	(N)	(19)	(28)	(18)	(65)
4	F	0.51	0.52	0.53	0.52
	D	0.35	0.37	0.38	0.37
	(N)	(13)	(13)	(15)	(41)
5	F	0.49	0.57	0.51	0.52
	D	0.35	0.29	0.35	0.33
	(N)	(26)	(23)	(25)	(74)
PAPERS 1 and 2	F	0.50	0.54	0.53	0.53
ALL QUESTIONS	D	0.30	0.30	0.33	0.31
	(N)	(94)	(95)	(96)	(285)

Implications of the Findings from Table 3.6

1. Those questions which have been classified as testing practical activities (Activity 2) have significantly lower discriminating values compared with the rest of the questions; that is, they are less likely to rank the candidates in the same way as does the examination as a whole. The mean discrimination index for questions devoted to practical work is 0.22 and that of all questions in Papers 1 and 2 is 0.31; this being significantly lower at the 1% level (T - test). It is also significantly lower than the mean value of D for all the other groups of questions in this classification (S - test 1% level).

This is of particular importance in this examination because these questions form a sub-test, the scores on which are used to moderate the teachers' assessment of practical work. It is necessary therefore, to examine the possible reasons for this difference in discriminating values:

(i) It could be that this sub-set of questions tests a particular homogeneous attribute of the candidates which is not tested, or not tested to the same extent, in the remainder of the questions. The particular attribute could be related to attainment in practical work, but there is no evidence to show that it is.

(ii) It could be that the sub-set of questions tests a heterogeneous mixture of attributes which are different from those tested in the remainder of the questions. Again, those attributes may or may not be related to attainment in practical work.

(iii) The reason may rest in the comparative unreliability of the sub-set of questions compared with the remainder, in the sense that candidates' scores on the same questions, if repeated after an interval, would not correlate highly with their original scores, even assuming that they had no memory of their original responses. The reliability of this sub-set of questions quoted in Appendix 6 arises from a measure of internal consistency (Cronbach's alpha coefficient) and is not necessarily related to the pre-test/post-test type of measure referred to above. Although the alpha coefficient is low, this may arise from the heterogeneous nature of the questions in the sub-set; that is, they may test many attributes not just one (see (ii) above). But again, there is no objective evidence to show that these attributes are, or are not, associated with practical work. All that can be said is that the sub-set of questions appears to be related to practical work and that candidates' performance in these questions differs from that in the remainder.

2. In contrast to Activity 2, those questions testing mathematical activities (4) have a significantly higher discriminating power. That is, compared with the rest of the questions they are more likely to rank the candidates in the same way as does the examination as a whole. The mean value of the discrimination indices of Activity 4 questions in 0.37 compared with the mean value for all questions in Papers 1 and 2 of 0.31 and of 0.22 for Activity 2 questions. The differences are significant at the 1% level. At the same time, it should be noted that the facilities of these groups of questions are very close to each other; what differences there are are not significant. We are concerned, therefore, with the fact that Activity 2 questions arrange candidates in an order of merit which is different from that of the other parts of the examination.

In addition, it is interesting to note that Activity 5 questions (using concepts not involving calculation) behave very close to the norm. While it must be acknowledged that this category of question is not so easy to define as those of Activities 2 and 4, it is worthy of note that it is the mathematical aspects of the use of concepts which gives rise to the characteristic type of discrimination in this examination.

This finding should be linked with those in the classification by Topic and Theme and serves as further evidence that those categories of question which are mathematical and physical behave more like the examination as a whole than the descriptive and non-physical categories. It seems logical to conclude therefore, that the examination as a whole has a mathematical and physical bias.

What action, if any, should arise from these findings?

1. Revision of the curriculum could alter the balance between the mathematical and physical areas on the one hand and the non-mathematical, descriptive and practical on the other. (This is not necessarily a distinction between conceptual and factual -see the comment about activity 5 above.)

2. The curriculum, in the form of the published materials, could be left unchanged, but the examination specification could be altered to give less weight to mathematical and physical categories of questions.

3. Both the curriculum and the examination could be left unchanged. This is tantamount to deciding that the present emphasis and manner of discrimination are the desired ones.

4. There are implications in those findings for the continued use of Activity 2 questions as a moderating instrument for the internal assessment of practical work. This will be discussed with other aspects of moderation in Section 5. which deals with the assessment of practical work.

COMPARISON OF ABILITIES

TABLE 3.7.

ABILITY		1970	1971	1972	Years Combined
1	F	0.55	0.50	0.55	0.53
	D	0.33	0.34	0.39	0.35
	(N)	(19)	(19)	(14)	(52)
2	F	0.52	0.58	0.57	0.56
	D	0.31	0.32	0.31	0.31
	(N)	(39)	(39)	(44)	(122)
3	F	0.48	0.47	0.49	0.48
	D	0.29	0.27	0.35	0.31
	(N)	(21)	(24)	(28)	(73)
4	F	0.44	0.59	0.43	0.49
	D	0.27	0.20	0.26	0.25
	(N)	(15)	(13)	(10)	(38)

PAPERS 1 and 2	F	0.50	0.54	0.53	0.53
	(D)	0.30	0.30	0.33	0.31
ALL QUESTIONS	(N)	(94)	(95)	(96)	(285)

Rather surprisingly, there proves to be little difference between the mean facilities in these questions designed to match the four categories of ability: knowledge, comprehension, application and analysis. The easiest questions are those in the comprehension category (2) and the most difficult those in the application category (3). The difference between the two is statistically significant at the 1% level (T - test), but the actual difference between the values (0.56 and 0.48), although fairly consistent over the three years, is not great and is not of an order which points to a need for change. Neither value is significantly different from the mean facility (0.53) of all the questions in Papers 1 and 2.

Whilst considering the facility values it is worthwhile pointing out that the trend from Ability 1 to Ability 4 is not significant, it follows that the examination questions as classified do not confirm a widely held hypothesis that questions in the higher categories of the Bloom taxonomy are necessarily more difficult than those in the lower categories. A further conclusion could be that the specification could be shifted from Ability 2 to Abilities 1, 3 or 4 without altering the mean facility of the examination as a whole.

Turning to the discrimination indices, it is rather surprising to note that those questions which most closely resemble the examination as a whole are those in category 1, knowledge (Mean D - 0.35); and those which least resemble the examination as a whole are in category 4, Application (Mean D - 0.25). Perhaps too much weight should not be put on the latter finding since consistent distinction between categories 3 and 4 is not easy. Nevertheless, it does seem, on the face of it, that the manner in which the examination as a whole discriminates between candidates resembles most closely the ability to recall. Further research may be necessary to prove this point conclusively, but it is of interest to note the direction in which the present findings point.

SUMMARY OF SECTION 3

Differing performance of candidates between one category of question to another is shown by differing mean facility values (F). In general it may be said that the causes could lie in one or more of the following:

1. The subject matter, or objective, a type of activity could be intrinsically easier or more difficult to learn.
2. The same areas may be intrinsically more difficult to teach.
3. The questions themselves are demanding too low or too high a standard of student ability.

The particular findings draw the attention of those responsible for the design of the curriculum and the examination to the following areas which may bear on decisions in the future:

1. Paper 2 discriminates between candidates better than does Paper 1. using the whole examination as a base. There is no significant difference between the facility values of the two papers.
2. In Paper 1 questions of the classification type have the

highest facility and discrimination; questions of the multiple completion type have the lowest facility.

3. Themes 2 (patterns in behaviour of materials) and 4 (kinetics) give rise to the highest performance by candidates: Themes 1 (specific behaviour of materials) and 5 (equilibria) give rise to the lowest performance.

4. Theme 1 questions give the lowest discrimination, Theme 5 questions the highest.

5. Topic 5 questions (halogens and oxidation states) give rise to the highest performance, Topic 9 (carbon chemistry I) the lowest.

6. Topic 15 questions (equilibria: redox and acid-base systems) show the highest discrimination. The discrimination of Topic 9 questions is significantly lower than that of the combined Paper 1 and 2 questions.

7. Activity 2 questions (practical work) show the lowest discrimination, Activity 4 questions (mathematical) the highest. There are no significant differences between the facilities of questions in the various categories of Activities.

8. There is little significant difference between the facilities of questions in the four categories of ability. Ability 1 questions (knowledge) show the highest discrimination, and Ability 4 questions (analysis) show the lowest.

9. In general it could be said that any increase in the weighting of the mathematical, physical, and recall sections of the examination and curriculum would reinforce the present type of discrimination to which the examination as a whole gives rise.

Any increase in Topic 9,(carbon chemistry 1) Theme 1 (specific behaviour of materials) and Activity 2 (questions on practical work) would tend to produce a different kind of discrimination than that which operates at present.

Finally, it must be emphasised that the findings give a factual basis for decisions; they do not make the decisions. If an area of the curriculum is showing aberrant behaviour in the facility or discrimination of the relevant responses of the candidate in the examination, this is not in itself sufficient reason for change. But the findings act as a monitor, drawing attention to those areas which do or do not conform to the norm and indicating the effect which any change would have.

SECTION 4.

THE FREE RESPONSE PAPER

Reference has been made already (Section 2.) to the need for a part of the examination in which candidates can answer under as little constraint as possible, a part in which there is a wide choice and which is subject to marking by impression rather than by an exact and detailed mark scheme. This is the function of Paper 3B (now Paper I) in which candidates are required to answer three questions out of a total of nine.

Many of the problems which arise when attempts are made to analyse performance on free response questions of which candidates have a choice are discussed by Nuttall and Wilmott in their book British Examinations.[7] Two main problems are the subjective nature of the marking with its consequent effect on reliability, and the element of candidate choice from which it follows that each question is attempted by a different group of candidates who may not be representative of candidates as a whole. Despite these two difficulties it was felt that an investigation into the Nuffield examination could not be complete without an analysis of the Free Response Paper, an account of which now follows.

The data were provided by the markers who, in addition to entering marks on the fly-leaf of the scripts, entered the marks on a special coding form (Appendix 4.). Thus for every answer two marks were recorded, one from each of the two markers marking out of a maximum of 5. These data were transferred to punched cards and a computer programme devised to yield the following information, which is summarized in Table 4.1 (1970), Table 4.2 (1971) and Table 4.3 (1972). Each Table being divided into three parts:

Part A summarizes the following information for each question in the Paper:

Number of attempts (N) at each question expressed as the number of answers received for a particular question and as a percentage of the total number of answers on the Paper.

Standard Deviation (SD) of all the marks (out of maximum 5) awarded by all the markers for a particular question.

Mean Score (MQ) of all the marks (out of maximum 5) awarded by all the markers for a particular question.

The Facility (F_A) of each question calculated by dividing the mean score awarded by individual markers (M_A) by the maximum score (5).

The Corrected Facility (F_C) of each question. This takes into account the fact that the group of candidates who attempted a particular question was not necessarily representative of all the candidates. The formula used for the correction was a modified form of the Willmott-Nuttall Facility Index (British Examinations p. 122):-

$$F_C = F_A + \frac{M_{GT} - M_{GTQ}}{100}$$

where F_A = Mean score on the question ÷ 5

M_{GT} = Mean percentage score on the whole examination by all candidates.

M_{GTQ} = Mean percentage score on the whole examination by that group of candidates who attempted the particular question.

The Discrimination (D) of each question, that is the correlation between performance on a particular question and performance on the examination as a whole. A similar correlation with performance on the Paper as a whole was also calculated, but has not been tabulated because each question comprises one third of the performance on the Paper, thus giving rise to a considerable bias in favour of a high correlation. There is a similar bias in the correlation with the examination as a whole, but it is very much less.

The Mean Score (M_{GTQ}) on the whole examination for that group of candidates who answered a particular question, expressed as a percentage.

Part B Summarises the following information:

The Number of Candidates (N) whose scores were analysed. In 1971 and 1972 this was nearly the total number of candidates who took the examination; in 1970 it was a sample of a little more than half of the total candidates.

The Standard Deviation (SD) of the scores (maximum 30) of all candidates on the Paper as a whole.

The Mean Score (M_p) of all candidates on the Paper as a whole (maximum 30)

The Discrimination Index (D) of the Paper as a whole, that is the correlation between performance on this Paper and the performance on the examination as a whole.

The Mean Score (M_{GT}) of all candidates on the examination as a whole, expressed as a percentage.

The Correlation between performance on this Paper and performance on other parts of the examination: The objective test (formerly Paper 1 now Paper 4) (P_4) the structured questions (formerly Paper 2 - now Paper 3) (P_3), the score for practical work (P_R) and the teachers' estimates of candidates' overall performance (T)

Part C summarizes the correlations for each question between the score awarded by one marker and the score awarded by another marker, and, in the last column, between total scores for each script.

TABLE 4.1 (1970)

Part A. Individual Questions

Q	N	SD	M_Q	F_A	F_C	D	M_{GTQ}
1	167 (15.8%)	1.06	1.91	0.38	0.40	0.47	50.3
2	131 (12.4%)	1.26	2.46	0.49	0.45	0.61	56.4
3	55 (5.2%)	1.15	2.24	0.45	0.47	0.62	49.9
4	193 (18.2%)	1.17	2.15	0.43	0.45	0.47	50.1
5	75 (7.1%)	1.30	2.13	0.43	0.39	0.55	55.4
6	52 (4.9%)	1.35	1.57	0.31	0.29	0.83	53.7
7	145 (13.7%)	1.24	2.04	0.41	0.37	0.62	55.9
8	73 (6.9%)	1.05	1.69	0.34	0.36	0.64	49.8
9	165 (15.6%)	1.10	2.07	0.41	0.47	0.47	45.5

Part B Whole Paper

N	SD	MP	D	M_{GT}	Correlations			
					P_1	P_2	P_R	T
360	5.14	12.1	0.76	51.3	-	-	-	-

Part C Mark/Re-mark Correlations

Questions	1	2	3	4	5	6	7	8	9	TOTAL
Correlation	0.58	0.77	0.83	0.70	0.73	0.85	0.79	0.80	0.71	0.85

TABLE 4.2(1971)

Part A. Individual Questions

Q	N	SD	M_Q	F_A	F_C	D	M_{GTQ}
1	381 (11.9%)	1.07	2.64	0.53	0.54	0.58	53.6
2	137 (4.2%)	1.16	2.16	0.43	0.35	0.58	62.1
3	484 (15.0%)	1.13	2.67	0.53	0.49	0.65	58.5
4	247 (7.7%)	0.83	2.37	0.47	0.51	0.41	50.3
5	562 (17.5%)	1.06	2.62	0.52	0.50	0.60	56.8
6	678 (21.1%)	0.77	2.45	0.49	0.52	0.33	51.7
7	221 (7.0%)	1.10	2.51	0.50	0.46	0.71	58.4
8	127 (4.0%)	1.05	2.73	0.55	0.50	0.59	59.8
9	371 (11.6%)	0.88	2.42	0.48	0.53	0.42	50.0

Part B Whole Paper

N	SD	M_P	D	M_{GT}	Correlations			
					P_1	P_2	P_R	T
1079	4.26	15.1	0.76	54.6	0.539	0.586	0.478	0.514

Part C Mark/Re-mark Correlations

Question	1	2	3	4	5	6	7	8	9	TOTAL
Correlation	0.78	0.81	0.85	0.71	0.77	0.63	0.83	0.76	0.70	0.84

TABLE 4.3(1972)

Part A. Individual Questions

Q	No. of attempts	SD	M_Q	F_A	F_C	D	M_{GTQ}
1	904 (14.1%)	1.03	2.29	0.46	0.44	0.59	54.1
2	169 (2.6%)	1.22	2.31	0.46	0.41	0.56	57.8
3	658 (10.2%)	1.14	2.29	0.46	0.41	0.58	57.5
4	795 (12.3%)	1.20	2.39	0.48	0.47	0.51	53.2
5	1041 (16.2%)	1.07	2.33	0.47	0.50	0.36	50.0
6	1853 (28.8%)	0.98	2.56	0.51	0.51	0.52	52.5
7	123 (1.9%)	1.16	1.80	0.36	0.34	0.60	54.3
8	208 (3.2%)	1.50	2.53	0.51	0.39	0.70	64.0
9	675 (10.5%)	0.92	1.90	0.38	0.45	0.30	45.5

Part B Whole Paper

N	SD	M_P	D	M_{GT}	Correlations			
					P_1	P_2	P_R	T
2174	4.46	1.38	0.75	52.5	0.536	0.558	0.463	-

Part C Mark/Re-mark Correlations

Question	1	2	3	4	5	6	7	8	9	TOTAL
Correlation	0.75	0.73	0.80	0.77	0.72	0.70	0.80	0.88	0.71	0.82

TABLE 4.4

CORRELATIONS BETWEEN VARIABLES ON FREE-RESPONSE PAPER

1970

	SD	M_Q	F_A	F_C	D	M_{GTQ}	Re-mark Correlation	Preference
SD	1.00							
M_Q	0.072	1.00						
F_A	0.071	0.994	1.00					
F_C	-0.407	0.818	0.809	1.00				
D	0.580	-0.479	-0.478	-0.716	1.00			
M_{GTQ}	0.761	0.189	0.203	-0.408	0.447	1.00		
Re-mark Correlation	0.409	-0.291	-0.283	-0.458	0.914	0.315	1.00	
Preference	-0.354	0.332	0.312	0.460	-0.756	-0.251	-0.777	1.00

1971

	SD	M_Q	F_A	F_C	D	M_{GTQ}	Re-mark Correlation	Preference
SD	1.00							
M_Q	0.198	1.00						
F_A	0.198	0.997	1.00					
F_C	-0.556	0.518	0.608	1.00				
D	0.928	0.379	0.371	-0.372	1.00			
M_{GTQ}	0.863	0.074	0.089	-0.736	0.776	1.00		
Re-mark Correlation	0.942	0.208	0.193	-0.480	0.946	0.755	1.00	
Preference	-0.402	0.248	0.220	0.505	-0.369	-0.440	-0.481	1.00

TABLE 4.4 (contd.)

CORRELATIONS BETWEEN VARIABLES ON FREE-RESPONSE PAPER

1972

	SD	M_Q	F_A	F_C	D	M_{GTQ}	Re-mark Correlation	Preference
SD	1.00							
M_Q	0.352	1.00						
F_A	0.366	1.00	1.00					
F_C	-0.520	0.499	0.495	1.00				
D	0.719	0.330	0.330	-0.593	1.00			
M_{GTQ}	0.884	0.473	0.478	-0.526	0.910	1.00		
Re-mark Correlation	0.856	0.010	0.122	-0.682	0.747	0.794	1.00	
Preference	-0.612	0.447	0.436	0.864	-0.363	-0.454	-0.596	1.00

Since there is only one instance of each of the eight variables in each year, high statistical significance cannot be attached to the correlations or lack of correlations which are shown. Nevertheless, some trends over the three years are worthy of note and can be summarised as follows:-

Standard Deviation and Corrected Facility

The correlations are negative in all three years indicating that these two variables are inversely related. Thus if the corrected facility of a question is low, it will tend to give rise to a greater spread of marks than if it were high.

Standard Deviation and Preference

It appears that these two are inversely related; unpopular questions tend to give rise to a greater spread of marks than do the popular questions.

Corrected Facility and Mean Score on Whole Examination

The consistent negative correlation over the three years indicates that those who do well in the whole examination tend to choose those questions which are awarded low marks, that is the most difficult questions.

Corrected Facility and Discrimination

Questions with a low corrected facility tend to rank the candidates in a similar manner tb the examination as a whole. The converse would be equally likely.

Corrected Facility and Re-Mark correlation.

Again, an inverse relationship. It appears that questions with a low corrected facility give rise to high correlation between markers, and vice versa.

Discrimination and Re-mark Correlations

These two variables show a consistently high correlation (0.914, 0.946 and 0.747), indicating that those questions which show high discrimination tend to give rise to high agreement between markers and vice versa.

Discrimination and Preference

An inverse relationship, indicating that the popular questions are less likely than are the less popular questions to discriminate between candidates in the same way as the whole examination.

Preference and Mean Score on Whole Examination

It appears that the least popular questions tend to be chosen by those who do well in the examination as a whole.

Preference and Re-Mark Correlations

The most popular questions tend to give rise to relatively low agreement between markers and vice versa.

CLASSIFICATION OF QUESTIONS

Unlike the objective and structured questions, the free response questions were not categorized in any way. Nevertheless, it is evident that free response questions are different from each other, so an attempt was made to determine whether particular types of question could be associated with one or more of the variables tabulated above. For each of the variables the three questions giving the highest rating and the three giving the lowest rating were listed (Tables 4.5 to 4.11) and the actual questions were examined to see whether they had some common characteristic which might influence the variables.

TABLE 4.5

ORDER OF PREFERENCE

1970

Order	Q	Content of Question
1	4	Practical work (film loop), variety of Topics
2	2	Topic 4, ionization energy, use of Books of Data
3	9	Theme 7
7	8	Topic 16, d - block elements
8	3	Background reading, halogens in fire-fighting
9	6	Problem on series of reactions
1971		
1	6	Periodic Table, use of Book of Data
2	5	Topic 8, structure
3	3	Topic 17, equilibrium and free energy
7	7	Topics 14 and 17.
8	2	Topic 12, use of Book of Data
9	8	Topic 16, d - block elements

TABLE 4.5 (contd.)

1972

Order	Q	Content of Question
1	6	Topic 4, use of Book of Data
2	5	Practical technique
3	1	Structure, Topics 8, 10 and 11.
7	8	Topic 12, heterogeneous equilibria
8	2	Topic 14, practical work
9	7	Equilibria, Topics 12 and 17

TABLE 4.6

ORDER OF STANDARD DEVIATION

1970

Order	Q	Content of Question
1	6	Problem on series of reactions
2	5	Practical problem. Topic 13.
3	2	Topic 4, ionization energy, use of Book of Data
7	9	Theme 7.
8	1	Topic 1, the mole and SI units
9	8	Topic 16, d - block elements.
1971		
1	2	Topic 12, use of Book of Data
2	3	Topic 17, equilibrium and free energy
3	7	Topics 14 and 17
7	9	Theme 7
8	4	Practical problem, organic chemistry
9	6	Periodic Table, use of Book of Data
1972		
1	8	Topic 12, heterogeneous equilibria
2	2	Topic 14, practical work
3	4	Topic 18
7	1	Structure, Topics, 8, 10 and 11.
8	6	Topic 4, use of Book of Data
9	9	Theme 7.

TABLE 4.7

ORDER OF QUESTION FACILITY (Actual)

1970

Order	Q	Content of Question
1	2	Topic 4, ionization energy, use of Book of Data
2	3	Background reading, halogens in fire-fighting
3	4	Practical work (film loop), variety of Topics
7	1	Topic 1, the mole and SI units.
8	8	Topic 16, d - block elements
9	6	Problem on series of reactions
1971		
1	8	Topic 16, d - block elements
2	3	Topic 17, equilibrium and free energy
3	1	Background reading, individual scientists.
7	9	Theme 7.
8	4	Practical problem, organic chemistry.
9	2	Topic 12, use of Book of Data
1972		
1	6	Topic 4, use of Book of Data
2	8	Topic 12, heterogeneous equilibria
3	4	Topic 18.
7	1/3	Structure / Topic 9
8	9	Theme 7
9	7	Equilibria Topics 12 and 17.

TABLE 4.8

ORDER OF QUESTION FACILITY (Corrected)

1970

Order	Q	Content of Question
1	3/9	Background reading, halogens in fire-fighting.
2	-	Sociological
3	2/4	Topic 4, ionization energy/Practical work (film loop)
7	7	Topic 15, use of Book of Data
8	8	Topic 16, d - block elements.
9	6	Problem on series of reactions.
1971		
1	1	Background reading, individual scientists
2	9	Theme 7
3	6	Periodic Table, use of Book of Data
7	3	Topic 17, equilibrium and free energy.
8	7	Topics 14 and 17.
9	2	Topic 12, use of Book of Data
1972		
1	6	Topic 4, use of Book of Data
2	5	Practical techniques.
3	4	Topic 18
7	2/3	Topic 14 / Topic 9
8	8	Topic 12,, heterogeneous equilibria.
9	7	Equilibria, Topics 12 and 17.

TABLE 4.9

ORDER OF QUESTION DISCRIMINATION

1970

Order	Q	Content of Question
1	6	Problem on series of reactions
2	8	Topic 16, d - block elements
3	3/7	Background reading /Topic 15.
7	-	Topic 1, the mole and SI units
8	-	Practical work (film loop)
9	1/4/9	Theme 7.
1971		
1	7	Topics 14 and 17
2	3	Topic 17, equilibrium and free energy
3	5	Topic 8, structure.
7	9	Theme 7
8	4	Practical problem, organic chemistry
9	6	Periodic Table, use of Book of Data
1972		
1	8	Topic 12, heterogeneous equilibria
2	7	Equilibria Topics 12 and 17
3	1	Structure, Topics 8, 10 and 11.
7	4	Topic 18
8	5	Practical techniques
9	9	Theme 7

TABLE 4.10

Order of Mean Score on Whole Examination of Group of Candidates who answered a Particular Question.

Order	Q	Content of Question
1970		
1	2	Topic 4, ionization energy, use of Book of Data
2	7	Topic 15, use of Book of Data
3	5	Topic 13, practical problem.
7	3	Background reading, halogens in firefighting
8	8	Topic 16, d - block elements
9	9	Theme 7
1971		
1	2	Topic 12, use of Book of Data
2	8	Topic 16, d - block elements
3	3	Topic 17, equilibrium and free energy.
7	6	Periodic Table, use of Book of Data
8	4	Practical problem, organic chemistry
9	9	Theme 7
1972		
1	8	Topic 12, heterogeneous equilibria
2	2	Topic 14, practical work
3	3	Topic 9
7	6	Topic 4, use of Book of Data
8	5	Topic 12, heterogeneous equilibria
9	9	Theme 7.

TABLE 4.11

ORDER OF RE-MARK CORRELATIONS

1970

Order	Q	Content of Question
1	6	Problem on series of reactions
2	3	Background reading, halogens in fire-fighting
3	8	Topic 16, d - block elements
7	9	Theme 7
8	4	Practical work (film loops)
9	1	Topic 1, the mole and S I units
1971		
1	3	Topic 17, equilibrium and free energy
2	7	Topics 14 and 17
3	2	Topic 12, use of Book of Data
7	4	Practical problem, organic chemistry
8	9	Theme 7.
9	6	Periodic Table, use of Book of Data
1972		
1	8	Topic 12, heterogeneous equilibria.
2=	3	Topic 9
2=	7	Equilibria, Topics 12 and 17
7	5	Practical Technique
8	9	Theme 7
9	6	Topic 4, use of Book of Data

Arising from Tables 4.5 to 4.11

It had been supposed that student preference for various parts of the course would be reflected when they exercised their choice in this Paper, but there is little in common amongst those questions high in preference or in those low in preference. Even when one applies other classifying principles such as descriptive/problem solving, open-ended and discursive/closed with a single correct answer, mathematical/non-mathematical, physical chemistry/inorganic and organic Chemistry, there is little in common at the top or bottom of the scales. A more detailed analysis by teaching sets might reveal that particular sets, either by preference or, as instructed by their teacher, show common trends in their preference; but for the examination population as a whole no such trends can be discovered.

An examinationof the Table of corrected facilities reveals that the factors which seem to produce questions associated with lowest performance are: physical chemistry, mathematics, and problem solving.

The factors which may be associated with high values of Facility are: descriptive, non-mathematical, discursive, with the exception of Q.6, 1972 and Q.2., 1970.

Conversely, Table 4.9 shows that factors associated with high discrimination are: physical chemistry, mathematics, problem solving. Significantly, Question 9 in all three years is a discursive question on sociological aspects of chemistry (Theme 7) and it always comes in the lowest group on the basis of discrimination.

Perhaps most significant of all is Table 4.10 which shows the performance in the whole examination of each group of candidates who chose each question. The Q.9 group always performed worst, indicating that there is a tendency for weak candidates to choose Q.9 in the hope of picking up marks while hiding their ignorance of chemistry. Whereas the questions chosen by the strong candidates tended to be physical chemistry, mathematical, and problem solving rather than discursive.

At this point Tables 4.5 to 4.11 should be considered together with the findings on pp. 32, 33 and 40. It appears that the better candidates not only tend to choose those questions comprising one or more of the physical, mathematical, problem solving factors, they also choose those questions with the lowest facility (corrected). Q.8 in 1972 is typical: it is concerned with physical chemistry (heterogeneous equilibria, Topic 12), it is mathematical, and it requires the ability to solve problems: the mean facility (actual) was 0.51, but if a typical group had answered rather than a high ability group, the corrected facility would have been 0.39. So, from one point of view, this group of candidates were penalized for choosing a difficult question. Another similar example is Q.2 in 1971.

These findings may reflect the fact that the questions in this Paper tend to fall into two broad categories; on the one hand, discursive, descriptive and imprecise, and on the other, problem solving, exact, and often mathematical. Coupled with this may be advice from teachers which could run something like this:
"If you know your chemistry, go for questions with a definite answer, if you don't, go for those on which you can waffle."

What action, if any, should be taken in view of these findings? There are grounds for changing the style of the Paper; for example, forcing candidates to do at least one question in each of the two broad categories mentioned above, or making all the questions into one style, or the other. But these may well be opposed on the grounds that candidates are constrained enough in the other Papers without constraining them in this, which, after all, is the only Paper in which there is a significant element of choice.

The problem of the corrected facility value is rather different. It may be that good candidates are being penalized by choosing questions which are marked severely and poor candidates are gaining unmerited marks by choosing questions which are being marked leniently. It would not be a difficult matter to correct marks satistically to take note of the different groups of candidates who answer the various questions, provided, of course, that the marks were coded in such a way as to facilitate rapid processing in time for the examiners' award meeting. But even this, apparently reasonable correction could be disputed on the grounds that the examination as a whole has a bias towards the physical, mathematical and precise, (see p. 45.) and that candidates with a broader view of chemistry should be allowed at least this opportunity of gaining some merit. The latter argument might well hold in this particular examination in which the free response questions have less than 20% weighting, but, if the weighting were ever increased, the need for a correction based on a corrected facility value would become more pressing.

GENERAL STATISTICS OF THE PAPER (Part B of Tables 4.1 to 4.3

There is not a great deal to comment on on the general statistics: The main factors do not vary much from year to year. The value of discrimination index is lower than that of the objective test and structured paper, but this is to be expected, even desirable in a Paper which has a substantially different function. Similarly, the correlations with the other measures, practical work and teachers' estimate, are lower, but consistent from year to year.

MARKING BY IMPRESSION

A summary of the correlations between the scores awarded by each of the two markers who read the answers to these questions is given in Part C of Tables 4.1 to 4.3 and a more detailed sample in Appendix 6 . In general it can be said that the correlations are sufficiently high on each question to reassure those who had doubts about this form of marking. The lowest correlations tend to apply to those questions giving rise to imprecise answers, for example Q.9. the sociological question; but even there, the correlations are sufficiently high to justify the continuance of impression marking. It will be noted that the correlation between two markers for the whole Paper is consistently high at above 0.80.

There are some qualifications which need to be made to the foregoing paragraph. The first is that a generally high correlation between two markers over a large number of scripts does not preclude the possibility of injustice to some individual candidates. It could happen that on a small number of occasions one or both markers could be abnormally severe or lenient and still achieve a high correlation over all scripts. If both markers are severe or both markers are lenient, the only way in which the anomaly can be detected would be to mark a third or more times, but, in view of the generally high agreement between markers, further marking may be regarded as wasteful of resources. The pairing of markers who are consistently severe or lenient over all scripts is a related problem.

The easiest form of abnormal marking to detect is that in which one marker is severe and the other lenient, giving rise to a large difference between the grades which they award. Provided the marks are recorded on appropriate coding forms, computer programmes can be devised which, in addition to the various statistics given earlier in this Section, will indicate occasions on which two markers differ by 2 or more grades on a particular answer. Arrangements can then be made for the answer to be marked by another examiner before the award meeting. As an illustration of the incidence of this sort of anomaly, in 1971, of 3208 answers which were marked, the two markers differed by 2 marks on 74 answers (2.3%) and by 3 marks on 3 answers (0.1%).

It must also be mentioned that the data on impression marking was gathered under operational conditions which were not as rigorous as could be applied to empirical research on impression marking. For example, although there were no marks recorded on the body of the answers, the marks of the first marker are recorded on the fly-leaf of the script and could have been seen by the second marker if the latter had wished to do so. This event is unlikely since the markers were specifically told not do do so; nevertheless, the possibility cannot be ruled out. It is also a fact that most of the markers were inexperienced in this form of marking and further experience could well lead to greater agreement between them.

As part of the normal marking procedure agreement trials are conducted before and during the marking of this Paper, by means of the common marking of a sample of photocopied scripts. These trials are of value in detecting severe and lenient markers and in reaching agreement on the criteria for marking. They are of less value in applying appropriate corrections; in fact it can be positively harmful to suggest to an examiner that he should mark more severely or more leniently, since there is a danger of over-correction.

There seem to be no grounds for grave concern over impression marking in the Nuffield Chemistry examination, but this is not to say that it could not be improved and the following suggestions may be thought worthy of further action.

(i) To make the coding and rapid processing of the marks a part of the routine procedure of the examination.

(ii) To introduce computer analysis of the agreement trials on common scripts so that the moderation of the marks awarded by examiners can be done statistically, rather than by word of mouth.

(iii) To introduce better sampling techniques, so that each pair of markers receives scripts from a group of centres representative of all centres. This will involve the monitoring of the performance of candidates from individual centres from year to year, a task which sounds more formidable than it actually is, provided suitable computer technology is used.

REFERENCES

7. NUTTALL, D and WILLMOTT, A. British Examinations: Techniques of Analysis. N.F.E.R. 1972.

SECTION 5.

THE ASSESSMENT OF PRACTICAL WORK

The Nuffield 'A' level Chemistry Project, in common with other curriculum projects of the same period, made a substantial departure from the traditional assessment by practical examination by adopting a form of internal assessment by teachers. Since this mode of assessment had been largely untried at G.C.E. level it was an obvious target for investigation. The essential elements of the assessment are given in Section 1. of this report and a fuller account is given in Appendix 2. (Readers who are unfamiliar with the Project are advised to read Appendix 2 before the remainder of this section.)

The main areas for concern were: the teachers' and students' perceptions of this form of assessment, the degree to which teachers followed a standard procedure for the assessment, and the operation of the moderating procedure. The first of these was investigated by questionnaire to teachers and students, the second by an analysis of the record cards of practical work, and the third by a comparison of the marks awarded in the practical examination with scores in other parts of the assessment.

THE TEACHERS' AND STUDENTS' PERCEPTION OF THE MODE OF ASSESSMENT

In this part of the work the Project was fortunate in having the assistance of Mr. Geoffrey Sleightholme[8], who developed the material in the form of a dissertation for the degree of M.A. The enquiry took the form of two postal questionnaires (Appendix 7. the first to chemistry teachers in all schools which entered candidates for the 1971 Nuffield examination; the second to all students entering for the 1972 examination in a sample of those schools. The sample for the student questionnaire was one-third of the total number of schools randomly selected, but keeping the proportion of girls, boys and mixed schools the same as in the full population. The numbers involved and the response rate is summarized in the following table

TABLE 5.1

Questionnaire	Type(s) of school	Number of schools approached	Number of schools replying	Total No. of questionnaires received	Response rates by schools
Teacher	All	53	49	99	92%
Student	Boys	10	9	Boys 208 Girls 57	90%
Student	Girls	3	3		100%
Student	Mixed	5	5		100%

FINDINGS FROM THE TEACHER QUESTIONNAIRE

The questions were in three groups: the first dealt with the principles and operation of the present system, the second with the effect of internal assessment on student/teacher relationships, and the third with possible changes in the system. Those who are interested in the detailed findings are referred to the Sleightholme dissertation; the following is a summary of the main points of interest.

The response rate to this questionnaire and that administered in the subsequent year (see Section 7.) was so high that it deserves special comment. Although the main hypothesis of the research is that examination outcomes can be used to monitor curriculum developments, it has never been supposed that examination outcomes were the only source of data. Direct monitoring of teacher and student opinion is of at least equal importance, and the two sources of data so complement each other that it was thought advisable to build the two questionnaires into the research programme. Looking to the future, it seems that any evaluation of curricula and examinations should include both techniques. With this in mind, the high response rate was encouraging and may be attributed to the following factors.

1. The questionnaires, particularly that of the students, were short and could be answered without prolonged thought. 2. The timing was such that it allowed teachers to complete the questionnaire either during the Easter holidays or early in the Summer term.

3. The questions concerned things about which teachers and students thought it important that their views should be known.

4. The teachers were given a summary of the findings and they knew that one of the objectives of the questionnaire was to use the findings, where appropriate, to improve both the curriculum and the examination.

In future work of this type the last factor is probably the most important of all. While recognising the useful work which teachers undertake on advisory and moderating committees of examining boards, one cannot escape the impression that the views of all teachers and students are not always adequately or accurately fed back into the evaluatory system. This investigation has shown that, keeping the factors listed above in mind and with proper sampling, the views of students and teachers can be systematically monitored. Teachers and students in the future may have to accept a certain amount of questionnaires as part of their professional and scholastic contribution to curriculum and examination development. Provided that it is self-evident that these views are used to change as well as monitor, there is every reason to believe that they will accept the small, additional work involved.

Teachers' opinions on the relative weighting of the five objectives of practical work showed that they were not entirely satisfied with the weightings at present operating. Their views are summarized in the following table:

TABLE 5.2

OBJECTIVE	TEACHERS' OVERALL OPINION ON WEIGHTING
1. Skill in observation	Slightly too high
2. Ability to interpret observations	Definitely too low
3. Ability to plan experiments	Rather low
4. Manipulative skills	Somewhat high
5. Attitudes to practical work	Somewhat high but not as high as (4)

The difficulty of making a meaningful assessment of attainment of each objective has been of some concern to the examiners. It seems, however, that teachers do not feel that such an assessment is difficult, with the exception of objective 3, ability to plan experiments, in which it appears that teachers would welcome guidance on how to assess it. This is shown in Table 5.3

TABLE 5.3

The Degree of Difficulty in Making a Meaningful Assessment

OBJECTIVE	Straightforward. % of responses	Difficult. % of responses	Very Difficult. % of responses
1. Skill in observation	60	33	5
2. Ability to interpret observations	68	28	4
3. Ability to plan experiments.	39	52	8
4. Manipulative skills	75	24	1
5. Attitudes to practical work.	67	27	6

The replies to Question 3 revealed that 83% of teachers normally assessed all the students during the same practical session and only 17% of teachers assessed some of the students. Rather surprisingly, Question 4 revealed that 48% of teachers frequently assessed students when the students were working in pairs; 35% did so infrequently; and 16% never did so. So a typical student, when having his work assessed, is likely to be working with a partner and all other members of his set will be assessed during the same practical session.

The 'Notes on Assessment' (Appendix 2,)claim that internal assessment by teachers could lead to "a greater awareness of the objectives of practical work". In an attempt to test this assertion, Question 7 asked teachers if they agreed that "the present system of assessment helps to sharpen the aims of practical work". The responses were:

Strongly agree	Agree	Not Sure	Disagree	Strongly Disagree
16%	56%	19%	9%	0%

which gives strength to argument that the involvement of teachers in assessment procedures does enhance their interest in the aims of their teaching.

One of the main problems confronting the members of a new curriculum project is in judging how much guidance to give to teachers and how much freedom to allow for individual judgement and initiative. Too much guidance and too detailed a structure given to an internal assessment system could defeat two of the main advantages of internal assessment, variety and flexibility.
On the other hand, complete freedom to assess what, when and how the teachers wish could lead to so diverse an assessment system as to be worthless for purposes of comparison. It is gratifying to note that the design of the system of assessment of practical work for this examination seems to have reached an acceptable compromise; 88% of teachers said that the form of assessment gave them sufficient freedom to assess as they thought appropriate. Of the 12% who did not agree, many complained that the specification of subject areas unnecessarily restricted them.

One matter for concern in the use of teachers in an assessment system is the demands made on their time, particularly on clerical tasks. In answer to Question 5, 14% of teachers thought the responsibility was not an extra burden, 58% that it was an extra burden falling within their normal duties, and 28% that it was an extra burden falling outside their normal duties. Although it seems clear that the Project has the goodwill of most teachers engaged in internal assessment, it must be noted that it is an additional burden and a substantial minority feel that remuneration should be considered.

When the internal assessment system was designed, it was acknowledged that it might change the relations between teachers and students and one of the main purposes of the questionnaire was to find out whether this was so (questions 9 to 13 in the teacher questionnaire and questions 4 and 5 in the student questionnaire).

In reply to the question "To what extent do you think student/teacher relations are affected in general....?", the teachers responded as follows:-

Distinctly improved	1%
Somewhat improved	8%
Not affected	89%
Somewhat impaired	1%
Distinctly impaired	1%

In reply to Question 11 74% of teachers said that they normally told their students of their intention to assess a particular experiment. Those who did tell their students were then asked the specific question, "To what extent have you found student/teacher relations affected when students are carrying out an experiment selected for assessment?" The responses were as follows:-

Distinctly improved	32%
Somewhat improved	17%
Not affected	52%
Somewhat impaired	7%
Distinctly impaired	1%

As to the effect on quality of practical work (Question 11.2) on occasions of assessment, the response was as follows:-

A marked improvement	4%
Some improvement	43%
No difference	24%
Some deterioration	1%
A marked deterioration	0%

It seems, therefore, that, in the teachers' view, their role as examiner did not have any general effect on relations with their pupils, and on specific occasions there was some feeling that relations and quality of work were improved. By the latter, teachers probably meant that the students were simply more diligent on occasions of assessment.

The last group of questions in the questionnaire to teachers was intended to obtain teachers' views on how, if at all, the system of internal assessment should be changed. Their general views on this form of assessment were expressed in answer to Question 14: 91% preferred some form of internal assessment, 6% preferred a practical test at the end of the course, and 3% preferred no assessment of practical work.

Those favouring internal assessment were then asked their general opinion of the particular form exemplified in Nuffield 'A' level chemistry. They replied:

Satisfactory as it stands	40%
Satisfactory but requiring minor alterations	48%
Satisfactory in principle but requiring major alterations	3%

Although a substantial number of teachers would like to see some change in the system, in the replies to the open ended questions 14 to 17 no distinct pattern could be found.

The suggestion of a single, general assessment by impression in the place of the present system was put forward in Question 18. The responses were:-

Strongly agree	12%
Agree	22%
Not sure	6%
Disagree	48%
Strongly disagree	11%

Although there is a wide distribution in these replies, the preference appears to be for the present system rather than the suggested alternative.

On the question of moderation of practical marks (Question 20) teachers were asked whether they agreed that "any form of internal assessment of practical work should, as a matter of principle be moderated by external means". The responses were:

Strongly agree	36%
Agree	46%
Not sure	9%
Disagree	5%
Strongly disagree	1%

From this response, it is clear that teachers would not be in favour of their raw scores for practical work being used unmoderated in the 'A' level assessment.

Nevertheless, in the space provided for general comment in the questionnaire a number of teachers felt strongly that the particular system of moderation used in this assessment is either unsatisfactory or obscure.

Concerning the communication of marks for practical work to students (Question 21) the response was:

After scaling and moderating YES 74% : NOT SURE 7% : NO 19%

Without scaling and moderating YES 59% : NOT SURE 10% : NO 28%

In general it seems that teachers are in favour of a fairly open system of communication of the marks obtained.

FINDINGS FROM THE STUDENT QUESTIONNAIRE

The general attitude of students to the assessment of their practical work was the subject of Question 1. The response to the statement "This form of assessment is preferable to a single practical examination at the end of the course" their response was: Strongly Agree 43%, Agree 45%, Uncertain 6% Disagree 4%, Strongly Disagree 2%. It must be acknowledged that the students were unlikely to have experience of the traditional practical examination in chemistry at this level, although they may have experienced one in other subjects. However, the evidence points to a large majority with a preference for internal assessment.

On the effect of internal assessment on their work (Questions 2 and 3) it seems that the general opinion of students is that it encourages greater care. To the statement, "This form of assessment encourages more care than might be the case" they responded: Strongly Agree 13%, Agree 51%, Uncertain 15% Disagree 18% Strongly Disagree 2%. And the assessment does not appear to give rise to undue anxiety: to the statement "this form of assessment is a source of anxiety in school work" they responded: Strongly Agree 5%, Agree 20%, Uncertain 18%, Disagree 45%, Strongly Disagree 12%. It must be remembered that the weighting of practical work in the whole assessment is only 15% and this must be a significant factor in the anxiety level in students arising from internal assessment. A related question (6) "this form of assessment tends to make one self-conscious in doing practical work" produced a similar response: Strongly Agree 5%, Agree 29%, Uncertain 13%, Disagree 43%, Strongly Disagree 10%.

Questions 4 and 5 are related to those in the teacher questionnaire which investigated the effect of internal assessment on the relations between students and teachers. To the statement, "this form of assessment has a tendency to change student/teacher relations for the better" the students responded: Strongly Agree 2% Agree 14%, Uncertain 46%, Disagree 31%, Strongly Disagree 6%. To the subsequent statement, "this form of assessment has a tendency to change student/teacher relations for the worse" they responded. Strongly Agree 0%, Agree 6%, Uncertain 35%, Disagree 48%, Strongly Disagree 9%.

In general, it seems that the students' opinion is much the same as that of the teachers: this particular form of internal assessment does not significantly change the relations between them. There is the further possibility that relations between student and student might be adversely affected, co-operation between students being a feature of most Sixth Form practical work. To question 8, "occasions of assessment in practical work reduce fruitful co-operation between students," students responded: Strongly Agree 5%, Agree 24%, Uncertain 12%, Disagree 44%, Strongly Disagree 14%. Although almost a third of the students do think that co-operation is reduced, the majority do not; so there seems to be little cause for concern. It is important to recognize, however, that responses to this question and to several others in the questionnaire might have been significantly different if the weighting of practical work were increased above the present 15%.

The objectives of the Nuffield practical work arose from consultation with teachers during the project trials, and with one exception (p. 74) it seems that teachers can recognize and assess achievement of the objectives. But it is perhaps equally important that students understand what the objectives are and what sort of performance is regarded as satisfactory evidence of the achievement of their objectives.

Answers to Question 7 indicate that this may not be so. The student response to the statement "this form of assessment helps to make clear the aims of each experiment" was: Strongly agree 6%, Agree 26%, Uncertain 21%, Disagree 35%, Strongly disagree 11%. From which it appears that students are less clear than the teachers of the sort of performance which constitutes a satisfactory achievement of an objective. Currently, there is much discussion about the need to inform teachers of the principles of formulation of objectives and assessment of achievement; perhaps it is equally important to inform the students.

THE PROCEDURES OF THE TEACHERS' ASSESSMENT

The two questionnaires reported above throw light on the opinions of teachers and students, but not much on the procedure by which teachers actually arrived at their assessment. The main source of the latter information comes from an analysis of the detailed record card which is completed by teachers for each student (Appendix 2.). Two analyses were done, one on the 1969 and 1970 record cards by Mr. L. Stringer[9], and the other by Mr. J.R. Leece on the 1971 record cards.

Average Number of Occasions of Practical Work on which Assessment Took Place, per Student.

In the notes for guidance for teachers a minimum of eight experiments per student during which assessment takes place is prescribed and this part of the research attempts to determine the degree to which teachers varied from the prescribed number. The findings are summarised in the following table.

TABLE 5.4

YEAR	No. of Sets	No. of Students	Average No. of occasions Assessment per student	Range of No. of occasions of Assessment.	% of sets with more than 8 occasions of Assessment.	% of Sets with less than 8 occasions of Assessment
1969	39	451	8.1	6 to 12	37%	33%
1970	60	659	9.5	6 to 16	57%	10%
1971	55*	575	9.9	8 to 18	71%	0%

* A random sample of the total population of 104 sets.

It can be seen that the average and the maximum number of occasions on which assessments take place per student has risen over the three years, indicating that teachers are fulfilling the intentions of the project in regarding eight experiments when assessment takes place as the minimum rather than the norm.

Frequency of Assessment of Each Objective

Table 5.5 summarises the number of times (percentages in brackets) a mark was given for the 1971 sample of candidates.

TABLE 5.5

Objective	No. of Assessments by Experiment	No. of Assessments by Impression over a period of time.	Total No. of Assessments
1	2105 (19.3)	76 (6.3)	2181 (18.0)
2	2840 (26.1)	16 (1.3)	2856 (23.6)
3	1832 (16.8)	21 (1.7)	1853 (15.6)
4	3466 (31.8)	416 (34.2)	3882 (32.1)
5	656 (6.0)	687 (56.5)	1343 (11.1)
TOTALS	10899 (100)	1216 (100)	12115 (100)

Number of Students = 575

Number of Experiments used as occasions of assessment = 5719

Total number of occasions of assessment (number of assessed experiments + number of impression assessments) = 5719 + 1216 = 6935.

This information indicates that there is little correlation between the frequency of assessment of each objective and the weighting attached to each objective. In particular, objective 5, although comprising 20% of the weighting was assessed much less frequently than the others.

It will be observed that the total number of times each objective was assessed by experiment (10899) is greater than the total number of occasions of experiment on which assessment took place (5719). This is because each experiment is normally used to make more than one assessment.

For each experiment used for assessment the average number of objectives assessed was 10899 ÷ 5719 = 1.9. This shows that the pattern established during the project trials, when it was agreed that two was the optimum number of assessments which could be made on each occasion of practical work, is being continued; but not universally so, because a range of 1 to 4 assessments on particular occasions was observed.

The number of assessments of each objective per student is given by dividing the numbers in Table 5.5 by the number of students (575) and is summarised in Table 5.6

TABLE 5.6

Objective	No. of assessments by experiment per student	No. of assessments by impression over a period per student	Total No. of assessments per student.
1	3.7	0.1	3.8
2	4.9	0.03	4.9
3	3.2	0.04	3.2
4	6.0	0.7	6.7
5	1.1	1.2	2.3
TOTALS	18.9	2.0	21.9

It follows that on average each student received approximately 22 scores from which his final mark for practical work was assessed. This, in itself, must contribute to the reliability of this form of assessment.

Marking Scales

During the project, it was found that a six point scale (5-4-3-2-1-0) was effective for most purposes. (In practice, it is a five point scale, since the zero is very rarely used). However, teachers were not obliged to use this scale, although it was recommended to them. The following table shows that teachers have continued to rely largely on this scale and in general find it to be the most useful.

TABLE 5.7

Year	% of Assessments		Other Scales
	Marked out of 5	Marked out of 10	
1969	92	6	2
1970	83	14	3
1971	90	9	1

Modes of Assessment

A general conclusion on modes of assessment for various objectives in practical work can be drawn. It appears that there are three modes:

1. Assessment on a particular occasion of practical work, using concrete evidence, usually written evidence. This mode is dominant for objectives 1, 2 and 3.

2. Assessment by impression on a single occasion of practical; the most frequent use of this being for objective 4.

3. Assessment by impression over a period of time, usually one term or more. This mode is used principally in the assessment of objective 5.

It can be seen from evidence already presented that the first mode predominates. Assessment by impression over a period of time is increasing, however. It comprised 6.7% of occasions of assessment in 1970 and 18% in 1971. A possible variation within the first mode is the use of experiments not included in the published course. The proportion of these experiments is not high, but it is increasing: 4.2% in 1969, 4.5% in 1970, and 8% in 1971.

The assessment of objective 4 provides a special case, in that all three modes are used and a comparison of the three modes becomes possible. Ten sets, comprising 100 students, were randomly selected from the 1970 population and the correlations between assessment for objective 4 by the three modes were calculated. These are shown in Table 5.8.

TABLE 5.8 CORRELATION MATRIX

4R = objective 4 marked from written records.

4I = objective 4 marked by impression on a single occasion.

4P = objective 4 marked by impression over a period of time.

	4R	4I	4P
4R	-	0.28	0.27
4I	0.28	-	0.68
4P	0.27	0.68	-

It is to be noted that, while the correlations between 4R and both 4I and 4P are low, the correlation between 4I and 4P is high This indicates the possibility that assessment of written records concerns a factor different from that assessed by impression. It also appears that both modes of impression marking of objective 4, whether on a single occasion or over a period of time, yield roughly the same outcome.

The Use of Particular Experiments for Assessment Purposes

In 1971 the nine most frequently assessed experiments were:-

1. Experiment 5.3 "To test predictions about the reaction between acidified solutions of potassium iodide and potassium iodate".
2. Experiment 7.2b "Evaluating an enthalpy change that cannot be measured directly".
3. Experiment 9.4b "The preparation of phenyl benzoate".

4. Experiment 16.1 "Investigation of oxidation states of vanadium".
5. Experiment 12.3 "To investigate the relationships between the concentrations of dissolved ions in equilibrium with a solid".
6. Experiment 10.2a "The approximate strength of the hydrogen bond".
7. Experiment 9.3c "The preparation of 1-bromobutane".
8. Experiment 13.2 "Identification of an aldehyde or a ketone".
9. Experiment 14.5 "The kinetics of the reaction between permanganate and oxalate ions.

The use of these experiments to test the various objectives is shown in Appendix 8, from which it can be seen that in seven out of the nine experiments one objective accounts for over 40% of the assessments. In five experiments this is objective 4, in one it is objective 2, and in one it is objective 3. It follows that, although most experiments can be used to test all the objectives, some experiments are regarded as particularly suitable for testing certain objectives; for example, Experiment 9.4b is predominantly (80%) used to test for objective 4.

STATISTICAL ANALYSIS OF SCORES FOR PRACTICAL WORK

An analysis was done of the scores awarded to all students in 1970 and 1971, using the following variables:

TABLE 5.9

VARIABLE	Possible Mark RANGE	ABBREVIATION
Weighted score in Objective 1.	0 - 25	OBJVE 1
do. 2.	0 - 15	OBJVE 2
do. 3.	0 - 10	OBJVE 3
do. 4.	0 - 30	OBJVE 4
do. 5.	0 - 20	OBJVE 5
Unmoderated Teachers Total	0 -100	TCHR TTL
Moderation Test Score	0-33(1970) 0-32 (1971)	MOD TTL
Moderated Practical Score	0 -100	ADJD TTL
Examination Grand Total	0 -100	SUBJ TTL

The mean scores, standard deviations, and correlation matrix for each of the years are shown below. The correlations were calculated for each set separately and then averaged. Sets containing fewer than 4 students were omitted.

TABLE 5.10

Means and Standard Deviations (1970)

MEANS	OBJVE. 1 16.745	OBJVE. 2 9.262	OBJVE. 3 6.362	OBJVE. 4 19.888	OBJVE. 5 13.351
MEANS	TCHR TTL 65.607	MOD TTL 17.102	ADJD TTL 69.037	SUBJ TTL 51.933	
STD DVNS	OBJVE. 1 3.164	OBJVE. 2 2.186	OBJVE. 3 1.520	OBJVE. 4 3.166	OBJVE. 5 3.028
STD DVNS	TCHR TTL 9.839	MOD TTL 4.610	ADJD TTL 13.224	SUBJ TTL 12.313	

TABLE 5.11

CORRELATION MATRIX (1970)

	OBJVE. 1	OBJVE. 2	OBJVE. 3	OBJVE. 4	OBJVE. 5
OBJVE. 1	0.999				
OBJVE. 2	0.514	0.999			
OBJVE. 3	0.447	0.452	0.999		
OBJVE. 4	0.472	0.449	0.345	0.999	
OBJVE. 5	0.474	0.530	0.488	0.568	0.999
TCHR TTL	0.788	0.756	0.651	0.783	0.814
MOD TTL	0.223	0.307	0.321	0.267	0.348
ADJD TTL	0.732	0.726	0.601	0.772	0.785
SUBJ TTL	0.432	0.565	0.459	0.405	0.529
	TCHR TTL	MOD TTL	ADJD TTL	SUBJ TTL	
TCHR TTL	0.999				
MOD TTL	0.374	0.999			
ADJD TTL	0.996	0.365	0.999		
SUBJ TTL	0.662	0.666	0.625	0.999	

TABLE 5.12

MEANS AND STANDARD DEVIATION (1971)

MEANS	OBJVE. 1	OBJVE. 2	OBJVE. 3	OBJVE. 4	OBJVE. 5
	16.304	9.095	6.196	19.467	13.605
MEANS	TCHR TTL	MOD TTL	ADJD TTL	SUBJ TTL	
	64.667	17.069	67.985	54.645	
STD DVNS	OBJVE. 1	OBJVE. 2	OBJVE. 3	OBJVE. 4	OBJVE. 5
	3.281	2.153	1.619	3.328	3.132
STD DVNS	TCHR TTL	MOD TTL	ADJD TTL	SUBJ TTL	
	10.786	4.145	13.092	11.737	

TABLE 5.13

CORRELATION MATRIX (1971)

	OBJVE. 1	OBJVE. 2	OBJVE. 3	OBJVE. 4	OBJVE. 5
OBJVE. 1	0.999				
OBJVE. 2	0.624	0.999			
OBJVE. 3	0.552	0.600	0.999		
OBJVE. 4	0.500	0.522	0.478	0.999	
OBJVE. 5	0.559	0.550	0.548	0.558	0.999
TCHR TTL	0.830	0.797	0.746	0.795	0.826
MOD TTL	0.206	0.282	0.284	0.252	0.269
ADJD TTL	0.810	0.792	0.739	0.775	0.801
SUBJ TTL	0.578	0.666	0.594	0.544	0.595
	TCHR TTL	MOD TTL	ADJD TTL	SUBJ TTL	
TCHR TTL	0.999				
MOD TTL	0.305	0.999			
ADJD TTL	0.996	0.319	0.999		
SUBJ TTL	0.737	0.551	0.751	0.999	

TABLE 5.14

CORRELATION MATRIX (1972)

	OBJVE. 1	OBJVE. 2	OBJVE. 3	OBJVE. 4	OBJVE. 5
OBJVE. 1	0.999				
OBJVE. 2	0.656	0.999			
OBJVE. 3	0.541	0.554	0.999		
OBJVE. 4	0.524	0.500	0.470	0.999	
OBJVE. 5	0.592	0.500	0.529	0.582	0.999
TCHR TTL	0.846	0.796	0.720	0.802	0.833
MOD TTL	0.328	0.310	0.262	0.257	0.307
ADJD TTL	0.826	0.783	0.710	0.789	0.822
SUBJ TTL	0.589	0.599	0.511	0.485	0.594

	TCHR TTL	MOD TTL	ADJD TTL	SUBJ TTL	
TCHR TTL	0.999				
MOD TTL	0.369	0.999			
ADJD TTL	0.997	0.382	0.999		
SUBJ TTL	0.710	0.655	0.712	0.999	

The first point arising from the analysis is that there is little difference between the means, standard deviations, and correlations between 1970 and 1971 (the figures are not available for 1972). The second is that the high mean and low standard deviation of the unmoderated scores awarded by teachers, which was first observed during the project trials, is maintained subsequently and appears to be a characteristic of this form of assessment.

The moderation of the teachers' scores by the use of the moderating score from the written papers has little effect on the mean sco

but

has a substantial effect on the standard deviation, resulting in a much greater spread of scores. This must arise from the scaling of each teacher's scores to one of three different means. Thus the bunching of scores which seems to be characteristic of a single set of candidates is not maintained in the scores of all the candidates.

The third point is to note the low correlation between the teachers' scores and the part score used in moderation, particularly in relation to the much higher correlation of the teachers' scores with the grand total for the examination. This is a matter of some importance, because an essential feature of a moderating test is that it should have a substantial common element with the moderated test. So, although there appeared a prima facie case for using written questions on practical work as the moderating instrument, the low correlation with the actual marks for practical work throws doubts on its suitability. If one is looking for a moderating instrument which correlates most highly with the teachers' scores, then the whole written examination, or at least the combined scores on the objective test and structured questions, would appear to be better, even if - on the face of it - they bear less resemblance to practical work. The low correlation with the part score may arise from the relatively low reliability of the part score compared with the written examination as a whole; nevertheless, the argument against its use as a moderating instrument stands.

The final point arising from the statistical evidence is the moderate and fairly uniform correlations between achievement in each of the five objectives of practical work; they also have a uniform correlation with the total examination score. This evidence points to the retention of the five objectives on the grounds that the correlations are low enough to suggest there may be sufficient difference between the objectives to warrant their retention. This, of course, is subject to calculations of the reliability of each of the sub-tests based on the objectives.

SUMMARY OF SECTION

A questionnaire on the opinions of teachers and students to the assessment of practical work met with a very high response. It is suggested that systematic monitoring of teacher and student opinion is both essential and feasible in curriculum development, provided the following factors are satisfied:

1. The questionnaires are short and can be answered without prolonged thought.
2. They are appropriately timed; for example, during the Easter holidays and the beginning of the Summer term.
3. The questions are confined to matters which are seen to be important to teachers and students.
4. The summary of findings is sent to all respondents.
5. Action is taken on the findings by the examination authority to improve both the curriculum and the examination.
6. In large scale examining standardised sampling and processing techniques are adopted.

Teachers indicated that they wished to see some changes in the weighting of the objectives of the assessment.

Teachers did not experience much difficulty in making an assessment with the exception of Objective 3, the ability to plan experiments.

Teachers tend to adopt standardised assessment procedures and to assess all members of their teaching set on the same occasion and on the same objectives. Variations in procedures tend to be between teaching sets and not within them. There is frequent assessment of students working in pairs. Most teachers agreed that the guidance given to them on the assessment was at the appropriate level.

Most teachers felt that their involvement in the assessment imposed an additional burden; and a substantial minority (28%) felt that the burden lay outside their normal duties.

Neither the teachers nor the students showed anxiety that the assessment worsened relations between them or caused undue anxiety. Teachers and students considered that, if anything, the assessment improved diligence and quality of work, particularly on particular occasions when students know that they were being assessed.

Both teachers and students favoured some form of internal assessment in general rather than a practical examination, but it must be noted that this group of teachers is not necessarily representative of all teachers. Of the Nuffield 'A' level assessment of practical work, 40% thought that it was satisfactory as it stands, 48%

satisfactory with minor changes, and only 3% thought that major changes were necessary. A substantial minority of teachers (34%) thought that a single general assessment by impression would suffice in preference to the present structured assessment.

Most teachers (82%) consider that moderation of the scores should be accepted as a matter of principle, but there were some doubts about the present system of moderation.

A substantial minority (29%) of students considered that the form of assessment reduced fruitful co-operation between them. This is not cause for concern at present, but it could be if the weighting for practical work were higher.

The number of occasions of assessment of practical work is rising from year to year and, now, a substantial number of teaching sets are assessed on more than the minimum of 8 occasions of practical work.

There is little correlation between the frequency of assessment of each objective and the weighting attached to it.

For each experiment used for assessment the average number ob objectives assessed was 1.9 with a range of 1 to 4.

The average number of scores for all objectives recorded for each student in 1971 was approximately 22.

On approximately 90% of assessments teachers prefer to use the 5-4-3-2-1-0 scale.

Assessment on a particular occasion of practical work using concrete evidence (usually written) was dominant for objectives 1, 2 and 3.

Assessment by impression on a single occasion was used most frequently for objective 4. Assessment by impression over a period of time was principally used for objective 5. The last form of assessment is increasing as is the use of experiments not in published materials.

The use of particular experiments for assessing particular objectives is shown on p.89 and Appendix 8.

Teachers' raw scores for practical work are characterised by high mean scores and low standard deviations. The effect of moderation is to increase the standard deviation, but it has little effect on the mean scores.

The correlation of the teachers' scores with the part scores used for moderation is low, but the correlation of teachers' scores with the total for the examination is high. This casts some doubts on the validity of the moderating instrument.

There is moderate and uniform correlation between scores for the five objectives, indicating that there is sufficient difference between the objectives to warrant their retention.

REFERENCES

8. SLEIGHTHOLME G. *Nuffield Chemistry: An Investigation into the Operation of Teacher Assessment of Practical Work at Advanced Level.* Unpublished M.A. Dissertation. University of Lancaster. 1972.

9. STRINGER, L. *Modes of Assessment of Practical Work in Science Curricula in Schools.* Unpublished M.A. Dissertation. University of Lancaster, 1971.

SECTION 6.

THE PERFORMANCE OF VARIOUS SUB-GROUPS OF NUFFIELD CHEMISTRY CANDIDATES

In looking for average performances in research of this type, there is a danger in assuming that the population under investigation is uniform, in this case collectively called Nuffield 'A' level chemists. But within this group lie several sub-groups whose performance might differ from that of the whole. The purpose of this section is to report on an investigation into the relative performance of these sub-groups in various parts of the examination.

The factors thought to be most likely concerned in variations in performance were: male/female, mathematicians/non-mathematicians, physicists/non-physicists, biologists/non-biologists. The subject categories were defined as those who did or did not take the subject at advanced level in the same year as they took Nuffield Chemistry. Two additional groups were those who did or did not take both maths and physics. The variables investigated were performance in the examination as a whole and in each of the separate papers of the examinations, including the teachers' assessment of practical work.

A summary of the findings is given in Tables 6.1. (Statistical significant differences are shown * (5% level) and ** (1% level).)

TABLE 6.1 MEAN SCORES (1970)

STUDENT SUB-SET	N	Paper 1	Paper 2	† Paper 3	Teachers' Practical Mark	Grand Total
Male	505	25.1	51.8	17.0	65.3	51.6
Female	91	24.4	53.4	20.4**	67.4*	53.8
Mathematician	340	26.3**	54.7**	17.9	67.1**	54.0**
Non-mathematician	256	23.3**	48.6**	17.0	63.7**	49.2**
Physicist	469	25.8*	53.9**	17.9	66.1	53.4*
Non-physicist	127	22.1**	45.3**	16.2*	63.8*	46.6**
Biologist	203	23.3**	48.7**	17.5	63.8*	49.7**
Non-biologist	393	25.9*	53.7*	17.5	66.5	53.1*
Both Maths and Physics	281	26.9**	55.9**	18.0	67.5**	54.8**
Neither Maths nor Physics	68	20.6**	42.1**	15.3**	62.9*	43.8**
All Students	596	25.0	52.0	17.5	65.6	51.9
Maximum Raw Scores		50	100	40	100	100

† Separate Scores for Papers 3A and B were not available.

TABLE 6.2 MEAN SCORES (1971)

STUDENT SUB-SET	N	Paper 1	Paper 2	Paper 3A	Paper 3B	Teachers' Practical Mark	Grand Total
Male	904	25.4	56.3	8.9	14.9	64.4	54.4
Female	142	25.8	57.1	9.9**	16.2**	66.5*	56.0
Mathematicians	591	26.8**	59.7**	9.0	15.3	65.9**	56.9**
Non-mathematicians	455	23.7**	52.1**	9.1	14.7	63.1**	51.7**
Physicists	749	26.5**	59.6**	9.3	15.6**	65.5*	57.0**
Non-physicists	297	22.7**	48.3**	8.6*	13.7**	62.5**	48.8**
Biologists	371	23.6**	52.3**	9.3	14.9	63.5*	52.0**
Non-Biologists	675	26.5**	58.6**	8.9	15.1	65.3	56.1**
Both Maths and Physics.	484	27.4**	61.5**	9.1	15.6**	66.3**	58.2**
Neither Maths nor Physics	190	21.9**	46.3**	8.4*	13.6**	61.7**	47.4**
All Students	1046	25.4	56.4	9.1	15.0	64.7	54.6
Maximum Raw Scores		50	100	20	30	100	100

TABLE 6.3 MEAN SCORES (1972)

STUDENT SUB-SET	N	Paper 1	Paper 2	Paper 3A	Paper 3B	Teachers' Practical Mark	Grand Total
Male	629	27.6	48.9	8.8	13.5	65.0	52.5
Female	104	27.7	47.6	9.9**	15.3**	66.4	53.6
Mathematicians	424	29.0**	51.8**	9.0	13.8	66.5**	54.7**
Non-mathemat--icians	309	25.7**	44.4**	9.0	13.6	63.3	49.7**
Physicists	482	28.7**	51.2**	9.1	14.0	66.2	54.4**
Non-physicists	251	25.5**	43.9**	8.7	13.3	63.3	49.1**
Biologists	283	26.3**	45.8**	9.4*	13.7	64.1	50.9*
Non-biologists	450	28.4*	50.6**	8.7	13.7	65.8	53.7
Both Maths and Physics	296	29.6**	53.1**	9.0	13.9	66.9**	55.7**
Neither Maths nor Physics	123	23.4**	38.8**	8.5	13.0	60.7**	45.6**
All Students	733	27.6	48.7	9.0	13.7	65.2	52.6
Maximum Raw Score		50	100	20	30	100	100

It is clear that those students who also take mathematics and/or physics do better in all parts of the examination, with the exception of Paper 3 than any other sub-group; conversely those who do not take mathematics and/or physics do worse.

Biology students do relatively badly, but whether this relates to taking biology or not taking mathematics or physics is difficult to determine.

These findings, together with those in Section 3. raise some interesting problems for the curriculum designers and examiners. Possible causes include:-

1. The study of mathematics and physics reinforces the study of Nuffield chemistry more strongly than does the study of biology and other subjects.
2. Those who study mathematics and/or physics have higher general ability than those who study other subjects and, therefore, tend to do better in any subject.

There is some evidence in Section 3. to support hypothesis but there is also evidence from elsewhere[10] to support 2. Whichever of the two predominates the findings have implications for the curriculum and examination designer. For example, if it were thought desirable to decrease the difference in performance between the subject sub-groups, the examination specification in the mathematical and physical dimensions could be decreased and the weighting on Paper 3 could be increased.

The difference in performance between boys and girls is consistent and worthy of note. In all three years girls have a higher mean total score for the examination as a whole than the boys, but the difference is not great. More significant is the fact that the difference in performance arises largely from a better performance by girls in both parts of Paper 3. This difference is quite large and is statistically significant at the 1% level. It follows that any increase in the weighting of Paper 3 would favour still higher performance by girls in the examination as a whole. Their performance in teacher assessed practical work is also slightly higher.

It could be argued that the girls are unrepresentative and of higher general ability than the boys. Against this is the fact that performance in Papers 1 and 2 shows no significant difference between boys and girls. It does seem reasonable to infer that there is something in the nature of Paper 3, particularly Paper 3B, which favours girls rather than boys.

The figures for performance in Paper 3A (1972 and 1971 only) show a small trend in favour of biologists. It has been reported to the examiners that Paper 3A favours biologists because the subject matter of the Food Science and Biochemistry Special Studies relates closely to the subject matter of 'A' level Biology, thus reinforcing their performance in these two Special Studies. However, it has already been shown that biologists do less well on the examination as a whole so, perhaps, their slight advantage in the Special Studies could stand.

Teaching Sub-Sets

During the research, the possibility became apparent of reporting on the performance of the particular groups of students who formed sets taught together. The possibility arises, almost by accident, from the mode of assessing the students' practical work, This is done on record cards by teaching sets (Section 5), thus permitting the coding of students by groups, each group being taught as a single unit.

Provided that recommendations on the coding and processing of marks, made elsewhere in this report, are accepted, it is now possible to report to each teacher the performance of his set of candidates in all the various elements of the curriculum compared with the national norm. Until now, the only feedback which teachers get is the examiners' general report, which refers to the population as a whole and not to individual teaching sets.

The further pursuit of this idea was outside the terms of reference of the research. Nevertheless, the usefulness of particular, objective, and quantitative feedback to teachers on their students' performance is self-evident and such feedback is now possible. There is further discussion of this important point in the last section of this report.

SUMMARY

This Section seeks to determine whether there are differences between the performance of sub-groups of the Nuffield Chemistry students.

Those students who also take mathematics and/or physics do better than other groups in all parts of the examination with the exception of Paper 3. Biologists do relatively badly. There are two possible causes of this:-

1. The study of mathematics and physics reinforces the study of Nuffield chemistry more strongly than the study of biology and other subjects does.
2. Those who study mathematics and/or physics have higher general ability than those who study other subjects.

Girls do slightly better than boys in the examination as a whole and in teacher assessed practical work, and significantly so in both parts of Paper 3.

Biologists do slightly better in Paper 3, particularly in the Special Studies, than they do in the rest of the examination.

REFERENCES

10. FORREST, G.M., SMITH, G.A., and BROWN M. Helen.
General Studies (Advanced) and Academic Aptitude. J.M.B. 1970, p. 19.

SECTION 7.

TEACHER AND STUDENT PERCEPTION OF SOME ASPECTS OF THE COURSE AND EXAMINATION

In Section 3 the areas of apparent difficulty and facility were determined by an objective analysis of the examination outcomes. What it did not reveal was the degree to which the actual outcome matched the teacher's and student's perception of the relative difficulty of the various parts of the course.

Since one of the main aims of the research is to establish techniques whereby the examination and examiners can become more closely related to the curriculum and teachers and students, it seemed important to determine the degree to which the actual outcome matched the expected outcome. To this end, two questionnaires (Appendix 7.) were administered, one to all the chemistry teachers teaching the Nuffield course in all the schools which took the Nuffield 'A' level Chemistry examination in 1972, and the other to all the students in a sample of 25 schools (16%) which took the 1972 examination. Replies to the teacher questionnaire were received from 143 schools (92%) Replies to the student questionnaire were received from 20 schools (80% of schools to which the questionnaire was sent and 13% of all schools).

The greater part of the questionnaire was devoted to the teachers' and students' perception of the facility of the Topics, it being felt that they would be insufficiently familiar with the other dimensions of the curriculum to make a judgement.

TEACHERS' AND STUDENTS' PERCEPTION OF TOPIC FACILITY

In Section B, Question2, of the teacher questionnaire, teachers were asked to indicate how difficult they thought their students found each topic by choosing one response on a five point scale. Thus a score of 1 indicated a very difficult topic and a score of 5 indicated a very easy topic. In order to allow a direct comparison with the Topic facilities calculated from the examination scripts, the scores were scaled to a maximum of 1. The scores on the student questionnaire were inverted, so that the easiest Topics were given the highest score, and again scaled to a maximum of 1 to allow a similar comparison. The results are shown in the following Table 7.1. (The number of teachers who did not make an estimate was insignifica for all Topics. The number of students who did not make an estimate was insignificant except for Topic 17, 90% response, Topic 18, 80% response, and Topic 19, 64% response).

TABLE 7.1

FACILITY VALUES BY TOPICS

TOPIC	TEACHERS' PERCEPTION	STUDENTS' PERCEPTION	1972 FACILITIES	COMBINED YEARS FACILITIES
1	0.73	0.86	-	-
2	0.60	0.67	0.47	0.52
3	0.52	0.59	0.45	0.50
4	0.43	0.51	0.59	0.57
5	0.54	0.52	0.68	0.68
6	0.55	0.47	0.49	0.49
7	0.38	0.45	0.60	0.59
8	0.33	0.42	0.59	0.55
9	0.46	0.42	0.44	0.46
10	0.53	0.49	0.09	0.55
11	0.56	0.46	0.30	0.43
12	0.31	0.33	0.55	0.50
13	0.46	0.39	0.52	0.52
14	0.35	0.35	0.71	0.58
15	0.29	0.31	0.33	0.46
16	0.49	0.44	0.49	0.44
17	0.17	0.25	0.60	0.50
18	0.51	0.47	0.70	0.55
19	0.52	0.48	0.55	0.63

Number of teachers	193
Number of students	257

It is interesting to note that the teachers' and students' perception of the relative facility of the Topics is very similar. This may be clearly seen in the scatter diagrams which were used to study the relationships between the questionnaire indices and examination facilities. Lack of space prevents the scatter diagrams being reproduced here, so the correlation coefficient is quoted as a convenient summary statistic. The actual correlation between teachers' and students' perception of the relative facility of the Topics is 0.87 and the two easiest Topics, 1 and 2, are identical and the three most difficult, 17, 15 and 12, are also identical. But there is much more difference between the teachers' and students' perception and the actual facility as shown in the examination outcome. The correlation matrix is shown in Table 7.2.

TABLE 7.2

CORRELATIONS OF FACILITIES

	TEACHERS	STUDENTS	1972	YEARS COMBINED
TEACHERS	1.00	0.87	-0.09	0.31
STUDENTS	0.87	1.00	0.09	0.48
1972	-0.09	0.09	1.00	0.54
YEARS COMBINED	0.31	0.48	0.54	1.00

(The 1972 figures should be disregarded because of the small number of examination questions in each Topic for a single year).

The Topic which proved to be significantly most difficult in the examination was Topic 9; whereas the teachers and students thought that Topic 17 was the most difficult. This raises some problems for the curriculum designer and for the examiner:-

Is Topic 17 intrinsically too difficult for students at this level or are both teachers and students mistaken?

Are the examiners compensating for the apparent difficulty of Topic 17 by setting questions which are trivial compared with the subject matter actually studied?

Either way, the research findings indicate that Topic 17 and, possibly, Topics 15 and 12, should be subject to special scrutiny by the examiners and curriculum developers as possible areas for change. Of course, it may be decided that no change is necessary; but it is a matter for some concern that both teachers and students find some topics so much more difficult than others; and it may be insufficient to assert that their anxiety is ill-founded simply because the performance on the examination questions on those Topics is not substantially different from the performance on the other Topics.

STUDENT INTEREST IN TOPICS

The first question in the student questionnaire was designed to determine the relative interest of the students in each Topic and Table 7.3 shows the mean responses scaled to an interest index in which maximum interest would score 100. To find out whether there was a significant difference between the interests of various categories of students, separate scores are recorded for boys and girls, and mathematicians and non-mathematicians.

The differences between the various categories of students are not significant and will not be pursued further; but there is a striking difference between the student's interest and perception of facility. The correlation between the two being negative (R= -0.44) Reference to Tables 7.1 and 7.3 shows that some of the Topics which were perceived as being very difficult (i.e. low facility) 17, 15 and 12, were not substantially below the average level of interest.

TABLE 7.3

STUDENT INTEREST IN TOPICS

TOPIC	ALL STUDENTS	BOYS	GIRLS	MATHEMATICIANS	NON MATHEMATICIANS
1	27	26	28	26	28
2	49	49	49	47	51
3	45	45	45	44	46
4	70	67	76	70	69
5	49	51	43	48	49
6	47	48	47	48	47
7	59	58	60	59	58
8	64	61	71	64	64
9	69	69	68	68	70
10	57	54	64	56	57
11	45	43	47	46	43
12	48	49	47	50	44
13	65	65	64	65	64
14	47	45	49	50	42
15	54	55	51	56	50
16	54	52	58	55	53
17	51	53	47	54	45
18	73	69	79	71	77
19	47	48	47	50	44
	N=257	N=176	N=81	N=160	N=96

Mean Interest Index for all students = $^{54}/100$ = 0.54

The three Topics considered to be of most interest were 18, 4 and 9 and yet all three were considered to be slightly more difficult than average.

The correlation between students' interest and actual facility was -0.25 which substantiates the evidence of the preceding paragraph: a high level of interest in a Topic does not necessarily produce a corresponding high level of performance in examination questions dealing with that Topic.

MISCELLANEOUS FINDINGS FROM TEACHER QUESTIONNAIRE

QUESTION 3

TABLE 7.4

TEACHERS' PERCEPTION OF NUFFIELD COURSE

	% responses (N = 193)
Should be completely re-written	1
Needs major changes	5
Satisfactory after minor changes	64
Satisfactory in its present form	16
Excellent in its present form	13

The Project team can take some satisfaction in these findings.

QUESTION 4

The question asked teachers to state the extent to which the examination constrained them in (a) what to teach and (b) how to teach.

TABLE 7.5

CONSTRAINTS OF THE EXAMINATION ON TEACHERS

(Responses given as % of teachers, N = 193.)

	WHAT TO TEACH	HOW TO TEACH
To a large extent	17	7
To some extent	33	31
To a small extent	25	27
Not significantly	25	35

The evaluation of Table 7.5. depends on one's views about the extent to which examinations should influence teaching. In general it seems true to say that the examination does not have such an influence on teaching as to give rise to anxiety.

Question 8

The question comparing experience of Nuffield 'A' level chemistry with that in traditional 'A' level chemistry gave rise to the following answers expressed as % response:-

"On the whole, WHICH WRITTEN EXAMINATION SYSTEM:	NUFFIELD (markedly)	NUFFIELD (slightly)	NO DIFFERENCE	TRADITIONAL (slightly)	TRADITIONAL (markedly)
(a) laid the greater stress on REMEMBERING FACTS	-	-	1	18	8
(b) laid the greater stress on UNDERSTANDING of CONCEPTS	74	25	1	-	-
(c) laid the greater stress on APPLYING experience in solving unfamiliar problems.	72	25	4	-	-
(d) tested the more COMPREHENSIVE selection of your students' work.	50	33	13	4	-
(e) produced the greater PRE-EXAMINATION TENSION amongst students.	-	7	63	25	6
(f) had the greater CONSTRAINING effect on your teaching.	6	23	22	30	20
(g) was easier for your students to pass.	7	43	36	12	1

SUMMARY OF SECTION 7.

The teachers' and students' perception of the relative facility of the Topics was very similar (R = 0.87), the two easiest Topics were thought to be 1 and 2,and the three most difficult 17, 15 and 12.

The correlation of the teachers' and students' perception with the actual facility was much lower (R - 0.31 and 0.48 respectively).

In the light of the findings, Topics 17, 15 and 12 should be subject to special scrutiny by the examiners and curriculum developers.

The students found that the three most interesting Topics were 18, 4 and 9, and the three least interesting were, 1, 3 and 11. The correlation between student interest and perception of facility and actual facility was negative (R = -0.44 and -0.25 respectively).

A large majority of teachers thought that the course was satisfactory, but required minor changes.

The examination appears to put some constraint on teaching, but not at a level to give rise to concern.

SECTION 8

CURRICULUM EVALUATION AND DEVELOPMENT BY CONTINUOUS PROCESSES

In the Introduction reference has been made to the main aim of the research, which was to investigate the possibility of using examination outcomes as an aid to a continuous process of curriculum development rather than a batch process. If continuous evaluation and development of a curriculum is to be achieved, a system of monitoring the operation of the curriculum is a prerequisite. To be complete, monitoring should be of two types. The first being intrinsic in which a particular curriculum system is monitored and evaluated without reference to other systems which impinge on it; the research reported in the foregoing sections was mainly of this type. The second is extrinsic in which the curriculum is related to other systems such as other curricula, further education, and future employment.

While recognizing that curriculum evaluation is not complete without relating it to external systems, the intrinsic evaluation is a necessary first-step. The purpose of this section is to summarize the conditions, constraints, processes, and possible outcomes relating to the continuous monitoring and development of a curriculum using the Nuffield Advanced Chemistry course and examination as an example. A general model of such a system is shown in Figure 1.

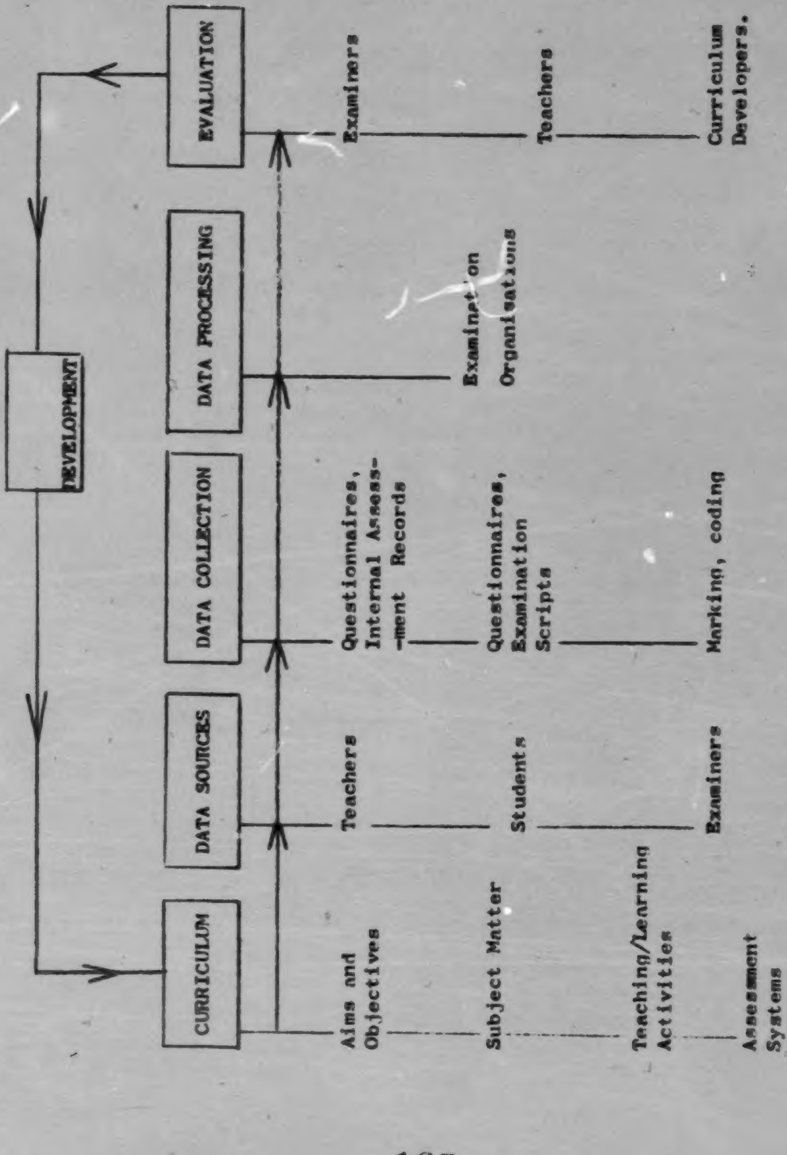

CURRICULUM
DATA SOURCES
DATA COLLECTION
DATA PROCESSING
EVALUATION
DEVELOPMENT
Aims and Objectives
Subject Matter
Teaching/Learning Activities
Assessment Systems
Teachers
Students
Examiners
Questionnaires, Internal Assessment Records
Questionnaires, Examination Scripts
Marking, coding
Examination Organisations
Examiners
Teachers
Curriculum Developers.

Figure 1.

CONDITIONS FOR CONTINUOUS CURRICULUM EVALUATION

Curriculum Structure

The curriculum, which is the object of the monitoring and evaluation system, is seen to have four main dimensions: The aims and objectives dimension in which the purpose of the curriculum is specified, the subject matter dimension, in which the content of the curriculum is specified, the activities dimension, in which the various types of teaching and learning processes are specified, and the assessment dimension in which the techniques by which attainment is measured, are specified.

The first condition for a continuous process of curriculum evaluation is that the specification of each of the first three is made explicitly and, where possible, quantitatively in the curriculum itself. There has been a move towards the fulfilment of this condition in recent years, particularly in science curricula. (For a general survey of this trend readers are referred to Schools Council Examinations Bulletin 27: Assessment of Attainment in Sixth Form Science). The structure for the Nuffield 'A' level curriculum was only partly specified in the published materials on the course itself. For example, the course specifies various topics for study, but does not indicate their relative weighting. Fortunately, the assessment procedures for the 'A' level examination were designed by those who had been involved in the design of the course. The examination specification (Section 1. of the report) therefore, can be taken to reflect the curriculum specification.

The Structure of the Assessment System

This leads to the second condition which is that the fourth dimension of the curriculum, the assessment system, should match the other three dimensions of the curriculum, both qualitatively and quantitatively.

This is not to say that the basis for the specification need be the same in all subjects. For example, the various skills and abilities which the students are expected to attain in the Nuffield 'A' level Chemistry course are classified in a similar manner to that of the Bloom Taxonomy of Educational Objectives ; whereas the skills and abilities promoted by the Schools Council Integrated Science Project are largely based on Gagne's Conditions of Learning . Whatever principle is adopted, it is a necessary condition for continuous monitoring that the structure of the curriculum in the three areas; objectives, content and process, is matched in the assessment system. Furthermore, it is important to accept that the fourth dimension, that of assessment, is an essential part of the curriculum and not something divorced from it. This applies both to the internal assessment operated by teachers and the external assessment by examination boards.

The Common Curriculum

Implied in the preceding paragraphs is the condition that a substantial part of a curriculum must be studied and assessed in common if continuous curriculum evaluation is to take place. In the Nuffield Advanced Chemistry course 60% of the assessment, the objective test and the structured question paper, allows no question choice to the students. Furthermore, all the main topics, themes, and student abilities and activities, are included in the assessment in approximately similar proportions from one year to the next. The actual proportion of a curriculum which should be common is a matter for debate. It is unlikely that a statistical treatment, similar to that described in the body of the report, would be worthwhile if the weighting allocated to the "common" curriculum fell much below 50%.

The Free Curriculum

It should not be thought, that continuous evaluation can be applied to only the common element in a curriculum. It is true that the evaluatory processes described in Section 3 of the report are designed for the common element only; but to limit curriculum evaluation to the common element would be to curtail both the range and the effectiveness of the evaluation. The Nuffield course encourages freedom of choice and approach in chemical education and this is reflected in the 40% "free" curriculum element in the final assessment, comprising the Special Studies, the free response questions, and practical work. Of course, the free element requires different forms of processing from those of the common element and will yield different information: In this element the relatively unconstrained choice and performance of students - and to some extent teachers - can be exercised. It is here that they can express their reaction to the curriculum and the examination, not through questionnaires or committees, but by what they actually choose to do or not to do.

Section 4 of the report describes one type of analysis which may be applied to that part of the Nuffield 'A' level written examination which allows a wide choice of questions and encourages a free response from the students. One of the conditions for a useful outcome of such an analysis is an appropriate system of classification of the questions. It is relatively easy to classify objective and structured questions according to the topic, ability and activity which they test, because each is designed to test a very limited area and to elicit a very short answer. But questions of the free response type give rise to much more complex and much longer answers which do not readily fit into the same classification. These questions are not classified in the present Nuffield examination, except - in a loose way - to describe them as tests of the ability to synthesize (Bloom level 5). It is acknowledged that the research did not give rise to an acceptable classification of these questions, except to recognize two main types: mathematical/physical/problem solving on the one hand, and discursive/descriptive/imprecise on the other. There is room for further work here.

Section 5 of the report describes the analysis of the other main part of the "free" curriculum, the assessment of practical work. Of course, no curriculum and its attendant assessment can be completely "free" if it is subject to the normal requirements of public examinations. The fact that the work is internally assessed, does not, in itself, bestow complete freedom to either students or teachers. In the Nuffield assessment of practical work, as in other forms of internal assessment, a compromise is reached between complete freedom and complete control by an examination board. If some comparability of work is to be achieved a common structure is necessary.

The form in which internally assessed work is described is not different in essence from that used earlier in this section to describe externally assessed work. The common element in the work has four dimensions: common areas of subject matter, common types of work, common objectives in terms of skills and abilities, and a common form of assessment. The essential condition to permit both freedom to teachers and students and continuous evaluation is to pitch the description of those four dimensions at such a level of generality as will allow sufficient common ground to make reporting meaningful and comparable together with sufficient freedom to prevent teachers and students from feeling constrained.

This is not to say that the structure of the curriculum is fixed, both qualitatively and quantitatively for all time. The whole point of continuous monitoring is to allow continuous evaluation and continuous adjustment,and periodic change of all the curricular dimension is almost inevitable. But, unless a general structure is provisionally agreed, monitoring and evaluation on a large scale cannot even start.

The foregoing paragraphs refer to the curriculum, not to standards of achievement. The latter, although based on the same common structure, introduce another factor, that of moderation of teachers' scores. Teachers of Nuffield 'A' level chemistry think that some form of moderation is necessary, but many have doubts about the moderation currently used. An evaluation of the moderating instrument is an essential part of the overall evaluation of internally assessed work.

Sources of Data

The emphasis so far has been on the need for a structure of the curriculum and assessment systems which has sufficient in common to allow continuous monitoring yet general enough to allow the necessary freedom to students, teachers and examiners. But the curriculum itself is inanimate and cannot speak for itself; the essential data can only come, directly or indirectly, from the three groups of people mentioned above. It follows that one of the conditions is that they must not only agree to a common structure, they must be able to understand it sufficiently to recognize and report on each part of it. For example, if teachers are to give marks for an objective such as "the ability to plan experiments," they must both agree that this is an appropriate objective and be able to recognize its achievement in their students. Similarly, examiners, students and teachers should be able to achieve substantial agreement on what distinguishes a question testing the ability to solve problems from one testing the ability to recall information. It is not sufficient that agreement is achieved amongst examiners and chief examiners. If reliable data is to come from the other two sources, they, too, must be able to recognize the essential dimensions of the curriculum and examination. So, a major condition for continuous evaluation is that teachers and students should understand the essential elements of the curriculum. It may be that they cannot all be involved in the original design; but, if this condition is satisfied, they all can be involved in its subsequent modification.

To students of systems analysis there is nothing new in this. All that is being said is that, if the curriculum system is to be effectively monitored and adjusted, information must be fed back from all the essential parts of it. But identifying the sort of information which is required and the sources from which it must come is only the first step; and in a curriculum with a well established structure like that of Nuffield chemistry it is comparatively easy. Much less easy, but no less essential, is the development of methods by which the data can be collected and processed.

THE COLLECTION AND PROCESSING OF DATA FOR CONTINUOUS EVALUATION

The three sources of data, students, teachers and examiners, have been already identified. Information from the written examination comes indirectly from the students and directly from the examiners; information on the practical assessment comes directly from the teachers. Teachers and students can contribute directly on several aspects of the curriculum and its assessment through the medium of questionnaires.

The Written Examination

Briefly, the information on the written examination comes from the responses to the fifty questions of the objective test, the scores on approximately fifty sub-questions of the structured paper, the scores on the special studies paper, and the scores on the free-response paper. The first is marked by machine, the second and third by detailed mark scheme, and the last by impression (twice).

There are three main aspects of the data arising from written questions and answers, their validity, their reliability and the techniques by which they are assembled and processed. The first, validity, has to some extent been dealt with earlier in this section when the point was made that students, teachers and examiners, should be agreed on what constitutes the essential structure of the curriculum and be able to agree on the classification of the various questions which test achievement in it.

Another aspect of validity is the degree to which the design of the examination allows the appropriate coverage of the curriculum, without which validity is not possible. This is particularly important in the "common" curriculum and is closely related to the number of questions in those parts of the examination which test achievement in it. The greater the number of questions the more likely is validity to be achieved. This is also a significant factor in the reliability of the assessment. In addition, the examination must be so designed that all candidates are required to attempt all the questions in those parts in which the "common" curriculum is covered.

The Reliability of each part of the examination can be estimated in various ways and an example of such estimates is given in Appendix 6 of the report. Clearly, these estimates of reliability should form part of the monitoring process. Reliability is a complex matter and cannot be discussed in full here. However, one aspect of reliability, internal consistency, impinges directly on any system of monitoring and requires special mention. (It was not possible in the research to investigate mark/re-mark reliability of individual examiners). A main contributory factor to internal consistency is the number of items contained within the test. The total number of items in the objective test and structured paper of the Nuffield examination is approximately one hundred. This is large compared with the more traditional forms of assessment and it contributes substantially to both validity and reliability.

The collection of data from the written examination.

This is in two parts: the first being the classification of the questions and the second the collection and coding of marks. The classification of each question in accordance with the main dimensions of the curriculum is an essential part of the process and the need for a system of classification which can be both understood and used by all those concerned with the course has already been mentioned. The Nuffield chemistry system in this respect was rather complex: subject matter for the written examination was classified under nineteen topics and seven themes; there were five different skills, five different types of activity and five different types of objective question. This is a very large number of variables and even with nearly one hundred compulsory questions each year it is not possible to include questions in all the cells of the resulting matrix. If the subsequent analysis of the statistics is to yield significant results, the relation between the number of questions and the number of variables is important. Under the present Nuffield system the amalgamation of scores on three years of examinations is probably necessary to yield significant findings. To obtain such findings after one year would require either more questions or fewer variables. It is not suggested that the structure of the course or the examination should be changed merely to facilitate the processing of data; nevertheless, this relationship between numbers of questions and variables must be kept in mind when planning continuous monitoring of this type. In the Nuffield course the nineteen topics represent convenient teaching units and on purely chemical criteria they are not all essentially different from each other; they could, therefore, be amalgamated into a smaller number of units to facilitate analysis without doing violence to the structure of the course.

In a subject like chemistry there seems no reason why a common system of classification of all questions in the "common" curriculum could not be agreed. This,accompanied by an agreed system of coding the various categories of question,could give rise not only to intrinsic evaluation such as that which has been attempted in this research but to comparative evaluation amongst all chemistry curricula at this level. Furthermore, it would then be possible to give precise statements about the attainment of students in terms of work covered and standards achieved. At present, the information available to those who rely on examination outcomes to guide students who are leaving school is far from precise.

The abstraction and coding of marks can be a time consuming and expensive part of the process. There is little problem with the objective test, which is machine marked and the data can be abstracted and processed within hours of the receipt of all the answer sheets. But in those parts of the examination which are marked by examiners the rate determining step is the transference by clerical assistants of marks from the scripts to coded mark sheets in such a form as to allow subsequent transfer to punched computer cards. The time and resources consumed by this process can be tolerated in a once-and-for-all research project, but not in a wider and routine monitoring process.

There are three possible solutions to this problem of data abstraction. The first is to apply the monitoring procedure solely to objective tests. It is suggested, however, that this solution should be rejected, because it would tend to raise objective testing to a level of importance which is unlikely to be acceptable by impairing the balance between fixed response and free response questions which is now one of the strengths of examinations in this country.

The second solution is to use sampling techniques, as this research project did in 1972, while retaining the present manual abstraction of marks. In a large scale examination, sampling could be at the level of one in five or possibly one in ten. However, two disadvantages remain: It would still take between three and six months to abstract and process the data and, in practice, this would involve a delay of twelve months before action could be taken on the findings. In long term curriculum development such a delay may not be important, but the research has shown that the data can reveal areas on which decisions should be taken before the publication of the examination results; for example in identifying disagreement between markers - particularly in the impression marking of the free-response part of the examination. Furthermore, rapid analysis of the answers to the structured questions could reveal the same sort of anomalies as does the item analysis of objective tests. Another defect in sampling is that the subsequent analysis can only reveal information about the whole population directly on that part which was sampled and indirectly on the whole population; information about the large group not in the sample would not be available. Thus, sampling would not permit feedback to individual schools and teaching sets which could be a feature of the monitoring system advocated in the report and discussed later in this section.

Probably the best way of overcoming the difficulties of collection and coding of marks is to require the markers to record the marks for each question and part-question directly on to a coding form. Since most examining boards require marks to be entered directly on to the scripts, it would be necessary to make the coding form an integral part of the script. An example of how this could be done for structured questions is shown in Appendix 5. By so doing, the marks could be transferred directly to a punched card, thus eliminating the expensive and time consuming abstraction of marks by clerical staff.

The Collection of Data on Internal Assessment

Marks for practical work in Nuffield chemistry are recorded by teachers on a separate record card for each student and on summary card for all the members of a teaching set. The coding of these marks is nearly as time consuming as that of the marks from the written examination. If continuous monitoring is to take place, a re-designed record card will be necessary so that teachers can record the marks in a form which is already coded. In this way, marks could be transferred directly to punched cards, in a similar manner to that suggested for the written papers.

The University of London School Examinations Department has already agreed to look into the feasibility of such a pre-coded system. If successful, this will not only facilitate the subsequent processing of the marks, it will reduce the large amount of rather tedious arithmetic which teachers have to undertake at present, and it will remove altogether the need for a summary card of the scores of each teaching set. Another advantage of pre-coded forms for internal assessment marks is that the date by which they are returned to the examining board could be set later because of the reduced time required for processing them. In their replies to questionnaires teachers have indicated a minority - but none-the-less significant - concern that internal assessment for public examinations is something extra to their normal duties: the more that can be done to streamline the process, and reduce the non-educational part of their task, the better.

Information from Questionnaires.

In most people, the prospect of having to complete a questionnaire normally produces reactions ranging from mild dislike to outright rejection; and teachers are no exception to this antipathy. But there is some information which can be collected in no other way. The actual marks which students obtain in the various parts of an examination, although they can reveal much information about a course and the examination itself, do not comprise a complete record of attainment nor do they reflect attitudes and opinions. The processing of marks is a fruitful and, so far, untapped source for curriculum evaluation, but it was felt that this research would be incomplete without complementary information obtained directly from teachers and students. Hence four questionnaires, two to teachers and two to students, were administered as part of the research and details are given in Sections 5 and 7.

The function of questionnaires in continuous curriculum evaluation is discussed in some detail in Section 5. In particular, the conditions necessary for the effectiveness of this part of the evaluation are set out on pages 72,73. It may be argued that the administration of such questionnaires is not part of the function of public examining boards; but, if the boards are to play a constructive part in curriculum development, they may be essential. Already examining boards tap teacher opinion through consultative committees and occasional conferences. The difficulty with the latter procedures is that they do not necessarily yield representative opinion, nor is the outcome readily quantifiable; high response to appropriately designed questionnaires can do both.

Information from Other Sources.

The foregoing passages in this Section relate solely to intrinsic curriculum evaluation. Although this was the main purpose of the research, one's attention was frequently drawn to the interaction with systems external to the particular curriculum under review; and if such a monitoring system were to be set up on a larger scale it would be desirable that data be gathered from other sources. In particular, the demands on the resources of an examining board of an assessment system could well form part of the investigation: so, too, could be the performance of the candidates in other subjects and, subsequently, in further education and employment. The research reported here did include some data external to Nuffield chemistry (see Section 6); but this is only a small part of what would be necessary for a comprehensive system of evaluation.

Data Processing

What has been written so far in this section deals with data collection. To be of use as a basis for decision, the information requires further processing. The details of this processing are too numerous and complex to be reported here. The main point to be made is that if continuous curriculum evaluation is to be effective and, above all - quick, full use of computerized techniques must be made. During the research many computer programmes were written to assist in putting the information into meaningful forms. These programmes will vary from examination to examination and their design can be demanding on computer programmers. But the difficulties are simply technical, and, provided that the information is suitably coded at the outset, lie well within the resources of examining boards. As a servant to continuous curriculum evaluation, the computer is a necessity.

A SERVICE TO TEACHERS

The information which arises from the processes described earlier in this section relates almost entirely to the average performance of all the students who study the Nuffield Advanced course. For curriculum development on a national scale it is this performance which must determine the judgements and possible changes which subsequently arise. However, during the research, it became apparent that the grouping of students into teaching sets for reporting attainment of practical work gave rise to the possibility of reporting on the performance of particular teaching sets in every aspect of the written examination.

Until now, the only feed-back teachers have had from the examiners has been in the form of general examiners' reports. These reports have tended to be subjective, non-quantitative and applicable to general performance, not the performance of particular teaching sets. If the type of collection and processing of data described in this report is adopted, it now becomes possible for the average performance of a particular group of pupils to be abstracted and sent to their teacher, together with the average performance of the examination population as a whole. In this way, a teacher could compare the performance of his group of students in all the various elements of the curriculum with that of the national norm **and to adjust his own curriculum if he thought it necessary.**

It is outside the terms of reference of this research to develop this idea further; but the usefulness of this information to teachers is clear and it could be obtained by relatively small changes in the data processing. Furthermore, the information could be obtained very rapidly - certainly in time for the start of the next school year. It is strongly recommended that, if continuous monitoring of some examinations is adopted, feed-back to individual schools should be an integral part of it **if teachers desire it.**

EVALUATION AND DEVELOPMENT

It would be wrong to think that a system of curriculum monitoring is merely mechanical. True, the area of data collection and processing forms a substantial part of the system and its design and operation is largely mechanical; but before and after this area lie two others in which the emphasis is more on personal judgement than mechanical processing. The first, the original design of the curriculum, has already been dealt with: the second, evaluation and subsequent development is discussed in this, the final part of the report.

The function of data collection and processing is to formulate problems, not to solve them. The findings arising from the research into the Nuffield Advanced chemistry course and examination provide evidence and identify areas in which change might be appropriate, not to make judgements and advocate changes; that function properly belongs to another body. Areas of the curriculum and examination on which decision and possible change might be desirable are identified in various parts of the report, some of the main ones being:

1. The structured questions have a significantly higher discriminating power than the objective questions, but they have similar facility values.

2. In the objective test, classification questions are both easier and more discriminating than the others.

3. There is little difference between performance in the various Topics, except that students perform higher than the norm in Topic 5 and lower than the norm in Topic 9.

4. There is little difference between the discriminating power of the Topics: Topic 15 has the highest and Topics 6, 9 and 10 the lowest.

5. Questions on practical work (Activity 2) in the written papers show a lower discrimination than any of the other questions. This is particularly important because it is this group of questions which is used to moderate the internal assessment of practical work. These questions also have low internal consistency.

6. Questions requiring mathematical skills (Activity 4) show the highest discrimination; yet there is no significant difference between the facility of the questions in the various categories of Activity.

7. Evidence from the objective test and structured questions suggests that an increase in the weighting of the mathematical, physical, and recall elements would reinforce the present type of discrimination exercised by the whole examination. Increases in Topic 9, Theme 1, and Activity 2 would have the reverse effect.

8. The difference between facility and corrected facility of some types of free response questions.

9. The high correlation between performance in the mathematical and physical aspects of chemistry and that in the course as a whole. And the high performance of those who take mathematics and physics at advanced level compared with those who do not.

10. The significantly higher performance of girls in the free response part of the examination.

11. The possible changes in the recording, collecting and processing

of marks, particularly for the structured and free response questions and the internal assessment of practical work.

12. The possibility of comparative evaluation amongst subjects, at least in science.

IN CONCLUSION

In general, it is probably fair to say that neither the Nuffield course nor its assessment system is malfunctioning in any serious way. The fact that any one part has been shown to function in a way different from the norm is not in itself sufficient reason for treating it as abnormal or for advocating change. Nevertheless, it would be surprising if a consideration of the evidence arising from the research led to a decision to make no changes at all. Perhaps the best example of the sort of questions a decision-making body might have to consider is the discussion on candidates' over-performance on Topic 5 of the Nuffield course and their underperformance on Topic 9, which is set out in Section 3.

The question arises, who should be responsible for evaluation, decision making and development, thus completing the cycle shown in Figure 1. It seems fairly obvious that the decision-making body should represent at least three main groups: the curriculum developers who originated the course, the teachers who have to teach it, and the examiners who have to assess attainment in it. Of course, the three groups are not necessarily distinct and it is encouraging that an increasing number of people exercise all three functions. Such bodies already exist in the various consultative committees of examining boards, the one relevant to this research being the Nuffield Chemistry Advisory Committee which is a sub-committee of the general Chemistry Advisory Committee of the University of London School Examinations Council.

It would be fair to say, however, that at present the main concern of these consultative committees is with monitoring and developing the examinations rather than the curriculum. But this is not a serious obstacle to a monitoring system which includes the development of both examinations and curricula; and there is a growing opinion that the two should not be regarded as separate. It may be thought to be a revolutionary step to integrate the development of curricula with the public examination system, but it certainly need not be a disruptive one. Whether it is desirable in all subjects is a matter for discussion; in science subjects, at least, there seems to be sufficient common ground between the various curricula and examinations to allow continuous monitoring and development. In fact, it may be more feasible in terms of use of resources if it is done on a large scale.

Many of the recent changes in curricula and examinations have been made intuitively and comparatively abruptly. There may be advantages in that sort of change. However, there is a case for having less abrupt changes and changes based less on intuition and more on objective evidence. It is hoped that this research report has indicated some guidelines on how a more evolutionary rather than revolutionary type of curriculum development could be adopted.

There is one danger in continuous curriculum development which must be mentioned: it is important that it does not degenerate into continuous curriculum tinkering. There is nothing more irritating to a teacher than a continuous stream of rather trivial changes in a curriculum and its examination; at all costs this must be avoided. It is essential that any changes must be important, generally agreed, based on objective evidence, and promulgated with sufficient notice. There is every reason for adopting continuous monitoring and evaluation; but it might well be advisable that any changes which arise from the evidence should be recurrent, possibly at three year intervals, rather than continuous.

REPORT ON RESEARCH INTO THE NUFFIELD ADVANCED CHEMISTRY COURSE AND EXAMINATION

APPENDICES

APPENDIX

1. Examples of Types of Questions used in the Nuffield 'A' level Chemistry Examinations.

2. Notes on the Assessment of Practical Work by Teachers (up to 1973).

3. An Example of Question Classification and Examination Specification.

4. Mark Sheets for Abstracting Scores.

5. Example of a Structured Question Designed for Pre-coding Marks.

6. Some Ancillary Data.

7. Questionnaires.

8. Measurement of Objectives of Practical Work in Conjunction with Particular Experiments.

APPENDIX 1.

Types of Questions Used in the Nuffield 'A' level Chemistry Examinations

MULTIPLE CHOICE

The manganate ion MnO_4^{2-} and the manganese (III) ion Mn^{3+} react separately in aqueous solution as follows:-

$$3MnO_4^{2-}(aq) + 2H_2O(l) \longrightarrow 2MnO_4^{-}(aq) + MnO_2(s) + 4OH^{-}(aq)$$

$$2Mn^{3+}(aq) + 6H_2O(l) \longrightarrow Mn^{2+}(aq) + MnO_2(s) + 4H_3O^{+}(aq)$$

These reactions resemble each other in that in both cases

A Water is acting as an oxidizing agent

B Water is acting as an acid

C The oxidation number of the manganese is increased

D The oxidation number of the manganese is both increased and decreased

E The oxidation number of the manganese is decreased.

CLASSIFICATION SET

Q. 1 to 5. These questions concern different types of forces between atoms and molecules. The headings A to E show five such types. For each of the questions which follow, select the heading which shows the force responsible for the effect stated.

A Hydrogen bonding
B Van der Waals' forces
C Ion-dipole attraction
D Dipole-dipole attraction
E Covalent linkage

Q.1. The surface tension of liquid tetrachloromethane (CCl_4).

Q.2. The formation of a compound between boron trifluoride (BF_3) and ammonia (NH_3).

Q.3. The provision of energy to break down the crystalline lattice of sodium chloride when it dissolves in water.

Q.4. The rigidity of candle wax (e.g. $C_{25}H_{52}$).

Q.5. The value for 120 found for the molecular mass of acetic acid (CH_3CO_2H, formula mass 60) in benzene solution.

SITUATION SET

Q 1 to 4.

When butan-1-ol is heated with excess potassium bromide and concentrated sulphuric acid, 1-bromobutane is formed. In one experiment starting from 3.7 g of butan-1-ol a 40 per cent yield was obtained. (Atomic masses: C,12;O,16;H,1;Br,80.)

$$CH_3CH_2CH_2CH_2OH \longrightarrow CH_3CH_2CH_2CH_2Br$$

Q.1. When butan-1-ol is changed into 1-bromobutane which of the following bonds must be broken?

A C - C

B C - H

C C - O

D O - H

E O - O

Q.2. In the formation of 1-bromobutane in this reaction which of the following bonds must be formed?

A C - Br

B H - Br

C Br - Br

D O - Br

E C - H

Q.3. Which of the following is LEAST likely to be a by-product of the reaction?

A but-1-ene, $CH_3CH_2CH=CH_2$

B hydrogen bromide, HBr

C but-2-ene, $CH_3CH=CHCH_3$

D potassium sulphate (or potassium hydrogen sulphate), K_2SO_4 (or $KHSO_4$)

E bromine

Q.4. Which of the following would be the yield of 1-bromobutane?

A 0.4 g

B 1.48 g

C 2.7 g

D 6.8 g

E 13.7 g

MULTIPLE COMPLETION

Directions summarized

A	B	C	D	E
1,2,3 only	1,3 only	2,4 only	4 only	Some other response or combination of responses

Q. Which of the following is/are an example of Gay Lussac's Law?

1. When the temperature of 20 cm^3 of hydrogen sulphide gas is raised from 300 K to 600 K its volume increases to 40 cm^3 (all measurements of volume made at 1 atm).

2. There are equal numbers of molecules in 20 cm^3 of hydrogen sulphide, 20 cm^3 of steam, and 20 cm^3 of sulphur dioxide (all measurements of volume made at 400 K and 1 atm.)

3. When the pressure on 20 cm^3 of hydrogen sulphide is decreased from 1 atm to $\frac{1}{2}$ atm, its volume increases to 40 cm^3 (all measurements of volume made at 400 K).

4. When 20 cm^3 of hydrogen sulphide gas is burned, it reacts with 30 cm^3 of oxygen to form 20 cm^3 of steam and 20 cm^3 of sulphur dioxide (all measurements of volume made at 400 K and 1 atm.)

RELATIONSHIP ANALYSIS

Directions summarized

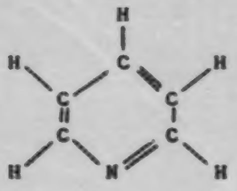
	Assertion	Reason	
A	true	true	Reason is a correct explanation
B	true	true	reason is not a correct explanation.
C	true	false	
D	true	true	
E	false	false	

	Assertion	Reason
Q.1.	The first ionization energy of helium is less than twice the first ionization energy of hydrogen.	The mutual repulsion between the two helium electrons partly offsets the attractive effect of the double positive charge on the helium nucleus.
Q.2.	In its compounds the oxidation state of potassium is NOT greater than one.	The first ionization energy of potassium is small and the second and subsequent ionization energies are much larger.
Q.3.	Some molecules do not react even if they collide with an energy greater than the activation energy for the reaction.	Some orientations of colliding molecules do not favour a reaction between them.
Q.4.	An experiment in which bromine (50 cm^3) is being transferred from one beaker to another should be done in a fume chamber.	If bromine is spilled on the fingers a painful burn is caused.

STRUCTURED

Q. 1. Cobalt(II) in aqueous solution can be determined by the following procedure:

i To the neutral solution add an excess of ammonium thiocyanate. Head to boiling and add an excess of pyridine. The formula of ammonium thiocyanate is NH_4CNS and that of pyridine is



ii Stir the solution and cool to room temperature. Red crystals are precipitated.

iii The crystals are filtered by means of a porous crucible which has been previously weighed. They are then washed with a solution of pyridine and ammonium thiocyanate in water followed with solutions of pyridine in alcohol and ether.

iv The crystals in the crucible are dried in a vacuum desiccator and then weighed.

v The weight of cobalt is calculated using the formula of the crystals:

$$[Co(C_5H_5N)_4]\ (CNS)_2$$

vi This determination of cobalt can be conducted in the presence of magnesium ions but not in the presence of copper ions.

Answer the questions which follow.

a What is the oxidation number of cobalt in the compound? 1

b Through which atom and by what type of bond will the pyridine be attached to the cobalt? 2

c Suggest a reason for the presence of pyridine in all the washing solutions. 2

d. What error might occur if ammonium thiocyanate were contained in the final washing solution? 2

e Suggest a reason why the presence of copper ions in the original solution interferes with this determination and magnesium ions do not. 3

FREE RESPONSE

Q.1. Use your Book of data to determine the lattice energy of strontium chloride. Show how you arrive at your answer.

A substance of formula $SrCl_3$ does not exist. What further energetics data would you need to explain this fact? Show how you would use the data.

Q.2. The Ancients knew of the existence of very few of the elements; by 1831, 47 were known and now we are aware of 103. Use your Book of data as well as your own knowledge of chemistry, to discuss the reasons why man did not discover all the elements at roughly the same time. In your answer, attempt to generalize about the nature of the elements discovered at different periods of history.

APPENDIX 2.

UNIVERSITY OF LONDON

(In association with all the G.C.E. Examining Boards)

GENERAL CERTIFICATE OF EDUCATION EXAMINATION

Nuffield Foundation Science Teaching Project

'A' LEVEL CHEMISTRY

Notes on the Assessment of Practical Work by Teachers

The purpose of these notes is to explain this form of assessment to teachers who are about to start the Nuffield 'A' level chemistry course.

The Advantages of Assessment by Teachers

1. The assessment can be done by those who actually observe the students at work over a long period of time, so that practical skills, abilities and aptitudes, which are not displayed in written work can be taken into account. In this way the objectives of the assessment can be made wider than those of a practical examination assessed on a single occasion on written evidence alone.

2. Assessment on several occasions is necessary to cover adequately the variety of experiments and techniques which comprise the practical work of the course; and the effect of chance failure or success is less.

3. The assessed practical work need not be limited to those experiments which can be readily administered to a large number of candidates in a limited time. The restriction of practical examinations to such experiments could have an undesirable effect on the variety of practical work in the schools taking the Nuffield course. The form of assessment set out in these notes should lead to a greater freedom of choice of practical work in schools.

4. This scheme does give teachers the opportunity to become involved in the assessment of their students. It could lead to a greater awareness of the objectives of practical work and a situation in which teachers and examiners work together towards these objectives.

The Outcome of the School Trials

During the two years of trials, eight experiments, all of them part of the basic course, were chosen by the Project's Headquarters team for assessment. Guidance was sent to teachers on the conduct of the experiments, the objectives of each, and the manner in which marks could be awarded. The results for each student were entered on a record card and the record cards were reviewed by the Examiners at the end of the first and second years. The main conclusions were as follows.

1. Teachers tended to use a restricted range of marks, and a candidate's total mark was rarely less than the half mark. The mean mark was correspondingly high. From an analysis of the marks, coupled with the direct observation of practical classes by the Headquarters Team, it can be inferred that the spread of ability is less in practical work than it is in written work, and it is unlikely that a student can be said to have failed in his two years of practical work.

2. The objectives of practical work were the subject of much debate. Finally, all the teachers who took part in the trials were invited to state what they considered the objectives to be. The objectives and their weighting, which are set out later in these notes, follow closely the outcome of this inquiry.

3. Many teachers felt that the detailed instructions, given by the Headquarters Team for each of the assessed experiments, unduly restricted them in their assessment. It became clear that no single set of instructions would meet the wishes of all the teachers, and it was decided that the proper course was to give as much freedom as possible to the teachers in the choice of experiments for assessment and in the conduct of the assessment. Some uniformity is necessary, however, in the objectives, the areas of work, the type of experiment, and the form in which the results are recorded.

4. It has been a constant policy of the trials that the assessment should be conducted on normal school work and be as unobtrusive as possible, but often there was a noticeable difference in the attitude of a class to the work when they knew it was to be assessed. Some teachers welcomed the more diligent work, others regretted the more formal atmosphere.

5. Teachers were rightly concerned lest the raw mark which they awarded their students would be the one used in the final 'A' level mark. The raw marks were used by the Examiners to determine only the distribution of marks within a teaching set, and this policy will continue.

6. Possibly the most difficult decision which had to be made was whether there was sufficient difference in the general standard of practical work between teaching sets and between schools to call for the use of a moderating instrument in order to make adjustments to the marks. For the time being, it seems that some form of moderation is desirable, but the matter will be reviewed from year to year.

7. Teachers found that those objectives which could be marked from written evidence were easier to assess than those which called for direct observation by the teacher during an experiment. It was found to be very difficult to assess more than one of the latter during any one experiment. It is hoped that assessment by direct observation will become easier as teachers become more familiar with it and can choose those experiments in which they know they can pick out particular faults and virtues without interfering with the normal teaching work.

THE FORM OF ASSESSMENT

The Objectives

The following objectives, with weighting given as a percentage, are suggested.

1. Skill in observation (25 %).
2. Ability to interpret observations (15 %).
3. Ability to plan experiments (10 %).
4. Manipulative skills (30 %).
5. Attitudes to practical work (20 %).

In many respects the form of assessment advocated in these notes has still to be evaluated, and it would be wrong to regard this statement of objectives and weighting as fixed for all time. Teachers may wish to modify both the objectives and the weighting, particularly in the finer divisions of the objectives. But for the time being it will help in the evaluation of this form of assessment if teachers can keep as close to the suggested pattern as possible. If teachers feel that they must deviate from this pattern, they will still be asked to show on the record cards how they allocated marks to particular objectives.

The objectives, 1-5, provide the broad basis for the final assessment, but for assessment on particular occasions a finer division is necessary. The following notes on the objectives may be helpful, but teachers should not feel that they are rigidly bound by them.

1. *Observation* alone has little merit unless it is accurate and relevant. Its assessment leans heavily on a student's ability to report, and it is inevitable that most of it will take the form of awarding marks for written records of observation. But there are two aspects of observation for which teachers can give credit while a conventional practical examination cannot, and it is hoped that neither will be overlooked. The first is oral reporting which is usually more immediate and less rehearsed than written reporting. The second is a student's observation and interest in unusual and unexpected features of an experiment even though they are not relevant to the immediate purpose of the experiment. It will not be easy to put a mark on these two aspects but at least credit could be given in a periodic review and grading of each student's attitude to practical work.

2. The assessment of *interpretation* is similar to that of observation in that teachers can only judge what students write or say. It will give credit for an analytical approach to observations, including unexpected observations and the ability to comment critically on practical situations.

3. The ability to *plan* is most easily assessed by asking the students to write their plan and submit it for marking before they start, but this can only be done in those experiments for which information to students on procedure is restricted. It should include the ability to recognise sources of error, and the ability to make predictions about other experiments.

It can be argued that both interpretation and planning can be assessed on written examination papers. This may be so, but a student's interpretation of observations which he has actually made and his plans for an experiment which he knows he must try to put into operation may well be different from his interpretation and planning of unreal situations on paper. The thinking processes may be similar but their context is different, therefore their outcome may be different.

4. *Manipulative skills* can be assessed by marking the results of experiments, for example yield, purity, and accuracy (accuracy of work rather than accuracy of calculation); they can also be assessed by direct observation of such things as orderliness, methodical working, dexterity and speed. It is desirable that this assessment should not be confined to routine operations; it should include the ability of students to adapt their manipulative skills to new situations and to follow unfamiliar instructions.

It may be possible for teachers to use specific experiments for assessing some parts of manipulative skill, particularly yield, purity and accuracy. However, many teachers feel that a grade based on a general impression of work over a period of say a term or even a year is more appropriate for the less exact aspects of manipulative skill which can only be assessed by direct observation. If periodic grading is used, it is important that teachers should consciously look for evidence of the skills throughout the period and not simply rely on their memory at the end of the period. (A further note on grading by impression is given on p. 3).

5. For the purpose of this assessment *attitudes* can be divided into three categories:

(i) Willingness to co-operate in the normal routine of a chemistry laboratory. This will include a knowledge of safety regulations and a willingness to comply with them and the other regulations which are necessary for the efficient running of a laboratory.

(ii) Persistence and resourcefulness—the will to work unaided on set problems and to see them through to a meaningful conclusion.

(iii) Enthusiasm; a commitment to practical work as a worthwhile pursuit without compulsion. This could show itself in suggestions for lines of investigation not specified in the course and in assuming the more active role when working in groups.

A grading at the end of each year, or even at the end of the course, is probably the most appropriate method of assessing attitudes.

Areas of Subject Matter

The assessment in the future will not be restricted to specified experiments as it was in the trials. Teachers will be free to choose which experiments to use and these experiments need not form part of the published Nuffield course. But in order to get a reliable order of merit it is important that the main areas of subject matter are covered. On each student's record card the examiners will expect to see a record of assessment on at least *eight* experiments, including at least *one* from each of the following areas of subject matter.

1. Changes in substances and patterns in changes in substances. (This is a deliberately broad category to include most of the course not covered by 2, 3, and 4, below. Most of the assessed experiments are likely to come in this category.)

2. Equilibria.

3. Kinetics.

4. Energetics.

Types of Work

Three main types of work can be identified in the course and the examiners will expect to see each one used in the assessment, but not necessarily with equal weighting:

1. Quantitative work.
2. Preparative work.
3. Qualitative work.

Allotment of Marks

The number of occasions on which marks are allotted for practical work is left to the discretion of the teacher provided that the minimum number of eight experiments is attained. It is desirable that the objectives, with the possible exception of objective 5, are assessed on more than one occasion in order to get a reliable estimate of ability.

Students should be told at the beginning of the course that their practical work will be assessed and that the results will be sent to the examiners. They should also know, in general terms, the qualities for which the teacher will look. They should be told that the assessment will be more obvious on some occasions than on others, but that it will usually take place during normal work and not on special occasions. It may reassure them to know that the average mark for practical work will be higher than that for the written examination.

The examiners will not necessarily expect that all the students in a teaching set will be assessed on the same experiments, but an equal balance of objectives is desirable. Assessment of objectives 1 and 2 will be difficult in experiments in which the students work in groups, but teachers should not feel restricted in the assessment to those experiments which are performed by students working alone: some assessment of objectives 3, 4 and 5 is possible in group work.

During the trials a grading scheme (5-4-3-2-1) was used to allocate marks to a particular objective on a particular occasion, and teachers may find a five point scale useful especially for those parts of the work which are marked by impression rather than by a mark scheme, but they are free to use other schemes if they wish.

There are three main ways in which marks can be allocated and teachers will probably find that they have to use all of them:

1. *A mark scheme*. This will be most useful when marking written evidence of observation, interpretation, planning, and accuracy.

2. *Marking by impression on a single occasion*. This will be useful for marking evidence which is less precise than that mentioned in 1. For example, a teacher may wish to assess dexterity in handling unfamiliar apparatus—say in chromatography. This can best be done by direct observation of the students at work, and in a large class, it is not easy. Teachers should try to assess only one such quality by impression during one experiment. Furthermore, teachers are usually too busy to do more than make a note of those who are doing well or badly; the rest can only be given the mark for average work.

A good deal of thought during the trials was given to the question of whether 'average' should be the teacher's estimate of the average for his set or his estimate of the average for 'A' level work in general. The former has the disadvantage that a teacher cannot estimate the general level of practical ability of his set until he has been teaching them for some time. The disadvantage of the latter is that it depends on the teacher's experience of 'A' level work. Since the raw marks will be scaled to a norm anyway, it is their distribution and the internal consistency of the marking which is important. But some uniformity is desirable and it has been decided that teachers should try to judge what is average from 'A' level work in general rather than from the work of their teaching set. Thus:

Excellence of work unusual in an 'A' level student	—	5
Above average work	—	4
Average work	—	3
Below average work	—	2
Unusually poor work for an 'A' level student	—	1

3. *Marking by impression after a period of time*. This will apply mainly to attitudes to practical work, but teachers may wish to assess other objectives, particularly some of the less precise aspects of manipulative skill, periodically rather than in single experiments. The period is left to the teacher's discretion; it could be done once a term, once a year, or, for some objectives, once at the end of the course. The criteria for awarding a grade are those set out in 2.

Information Required by the University of London

Three or four weeks before the written examination the School Examinations Department, University of London will require the following information from schools.

1. All the record cards of the students arranged according to teaching sets. A sample showing how a record card can be completed is given in Appendix A; this should not be taken to be an example of an ideal assessment and the experiment numbers refer to trial editions of the course.

2. A summary of the marks for practical work for each set, showing the total mark awarded to each candidate as a percentage (Appendix B); copies of this form will be sent to schools nearer the time of the examination.

Any other relevant information should be recorded on form C31 which will be distributed at the same time as the summary forms. For example:

(i) Outstandingly good or bad performance throughout the course by individual candidates.

(ii) Outstandingly good or bad performance by a set as a whole.

(iii) Circumstances such as absence or change of school which could affect the performance of a candidate, particularly if it has resulted in the completion of fewer than eight experiments in which assessment has taken place.

The Moderation of Marks

The present need for moderation of the teachers' marks has already been mentioned (p.1.). For various reasons, including those of ineffectiveness and expense, moderation by visiting examiners, moderation by a practical examination, and moderation by performance on the written papers as a whole, have been rejected.

Moderation will take place by the performance of teaching sets as a whole in those questions in the written Papers 1 and 2 which will have been designed to test practical work; these questions will form not less than 20% of the two Papers. It seems that performance in these questions is more likely to reflect the general level of practical work of a teaching set than any other moderating instrument which can be conveniently applied.

Three ranges of marks will be allocated, based on the performance of teaching sets as a whole in the written practical questions. To conform with the findings of the trials (p.1.) all three ranges of marks will be narrow, they will overlap, and all three mean marks will be higher than the half mark. For example, in 1968 the maximum mark for the teacher assessment of practical work was 70, which comprised 14% of the maximum mark for the whole examination, and the three mark ranges were:

Top range (20% of teaching sets), mean mark 57.5, standard deviation 4.

Middle range (60% of teaching sets), mean mark 47.5, standard deviation 4.

Bottom range (20% of teaching sets), mean mark 37.5, standard deviation 4.

This form of moderation will be reviewed from year to year.

In Conclusion

This is a first venture in this form of assessment of 'A' level practical work in chemistry by teachers, and in many ways it is still on trial. At first it will no doubt seem strange to both teachers and students, but it is hoped that in time it will become accepted as part of normal Sixth Form work.

Research into the operation of the assessment is starting in 1970. Queries, comments, and suggestions, are welcome and should be addressed to J. C. Mathews, Department of Educational Research, University of Lancaster, Bailrigg, Lancaster.

UNIVERSITY OF LONDON

(In association with all the G.C.E. Examining Boards)

GENERAL CERTIFICATE OF EDUCATION EXAMINATION

Nuffield Foundation Science Teaching Project

CHEMISTRY, ADVANCED LEVEL

Assessment of Practical Work

School Name .... Centre Number .... Name of Candidate .... Form and/or Set ....

Duration of course: September, 19 to July, 19

RECORD CARD NUMBER 1 of 2

ASSESSED EXPERIMENTS AND OTHER ASSESSMENTS		OBJECTIVES AND MARKS (SHOWING MAXIMUM)				
		1 OBSERVATION 25%	2 INTERPRETATION 15%	3 PLANNING 10%	4 MANIPULATION 30%	5 ATTITUDES 20%
2.2(c)	Properties of chlorides	6/10	4/5			
5.3	Iodide/Iodate reaction				7/10 (accuracy)	
6.1(b)	Reactions of some metallic elements	8/10	3/5			
7.2(b)	Enthalpy of hydration			3/5 (Comment on sources of error)		
9.4(b)	Preparation of phenyl benzoate				2/5 (yield) 4/5 (purity)	
General impression of manipulative skill (year end)					4/5 (speed) 3/5 (orderliness)	
General impression of attitudes to practical work (year end)						4/5 (routine) 4/5 (persistence)

SUMMARY

(Carried forward)	OBJECTIVES					TOTALS
	1	2	3	4	5	
MARKS						
SCALED MARKS						

....
Signature of teacher

UNIVERSITY OF LONDON

(In association with all the G.C.E. Examining Boards)

GENERAL CERTIFICATE OF EDUCATION EXAMINATION

Nuffield Foundation Science Teaching Project

CHEMISTRY, ADVANCED LEVEL

Assessment of Practical Work

School Name.................................... **Centre Number**.................... **Name of Candidate**.................................... **Form and/or Set**..............

Duration of course: September, 19 to July, 19

RECORD CARD NUMBER 2 of 2

ASSESSED EXPERIMENTS AND OTHER ASSESSMENTS		OBJECTIVES AND MARKS (SHOWING MAXIMUM)				
		1 OBSERVATION 25%	2 INTERPRETATION 15%	3 PLANNING 10%	4 MANIPULATION 30%	5 ATTITUDES 20%
10.2(a)	Strength of hydrogen bond		3/5	5/10		
12.3	Solubility product of silver iodate				4/5 (accuracy)	
14.2	Rate of reaction			4/5 (recognition of errors)		
17.1	Reaction of thionyl chloride with hydrates	4/5	3/5			
17.3(a)	Redox reactions	6/10		3/5 (prediction)		
18.3(a)	Chromatographic separation of amino acids	4/5	2/5		3/5 (dexterity)	
General impression (year end)					4/5 (speed) 4/5 (orderliness)	4/5 (routine) 4/5 (enthusiasm)

SUMMARY

	OBJECTIVES					TOTALS
	1	2	3	4	5	
MARKS*	28/40	15/25	15/25	35/50	16/20	109/160
SCALED* MARKS	18/25	9/15	6/10	21/30	16/20	70/100

....
Signature of teacher

*Note: Marks should be totalled and scaled at the end of the course only, do not total and scale marks for each year and give the average scaled mark.

15

APPENDIX B

UNIVERSITY OF LONDON

(In association with all the G.C.E. Examining Boards)

GENERAL CERTIFICATE OF EDUCATION EXAMINATION

Nuffield Foundation Science Teaching Project

CHEMISTRY, ADVANCED LEVEL

Summary of Assessment of Practical Work

School .... Centre Number* .... Form and/or Set ....

Duration of Course: September, 19 to July, 19

Candidate Number*	Name and Initials	Total Mark /100

* Insert the University of London Centre or Candidate Number for the examination; these should be known at the time this form is submitted, otherwise the relevant spaces may be left blank.

Note: A separate form should be used for each teaching set.

Teacher's signature .... Date ....

APPENDIX 3.

PAPER I, SAMPLE EXAMINATION SPECIFICATION

Item	Topic	Theme	Activity	Ability
1	14	4	D	1
2	4	3	C	2
3	14	4	E	2
4	1	2	D	2
5	9	1	A	2
6	8	3	C	2
7	16	2	D	2
8	7	6	D	2
9	16	3	E	3
10	9	1	A	2
11	12	5	D	2
12	12	5	D	2
13	6	6	D	3
14	15	5	D	3
15	9	1	B	4
16	16	2	D	2
17	17	5	D	3
18	13	1	C	2
19	18	1	C	1
20	13	1	C	1
21	9	1	C	1
22	8	3	C	2
23	8	3	E	2
24	8	3	E	3

Item	Topic	Theme	Activity	Ability
25	18	1	A	4
26	18	3	A	4
27	18	7	A	2
28	18	2	A	1
29	9	1	B	3
30	9	1	B	3
31	9	1	B	3
32	9	1	B	3
33	9	3	E	3
34	15	1	B	1
35	13	1	A	1
36	7	6	C	2
37	5	2	E	2
38	5	2	C	2
39	7	6	E	1
40	2	2	C	2
41	14	4	C	3
42	7	6	B	3
43	7	6	D	3
44	9	1	B	3
45	4	3	C	2
46	4	3	C	1
47	8	3	E	2
48	6	2	E	2
49	13	3	C	2
50	15	5	E	2

APPENDIX 3. (Continued)

PAPER 2 SAMPLE EXAMINATION SPECIFICATION

Item	Topic	Theme	Activity	Ability
1(a)	12	5	B	2
(b)	12	5	B	2
(c)	12	5	E	2
(d)	15	5	B	2
(e)	12	5	D	3
(f)	10	5	C	3
2(a)	16	3	E	2
(b)	16	3	E	1
(c)	16	3	E	3
(d)	16	3	E	3
(e)	16	2	E	3
(f)	16	3	E	4
(g)	11	3	E	4
3(a)	7	6	A	2
(b)	7	6	A	2
(c)	7	6	A	2
(d)	7	6	A	2
(e)	7	6	A	1
(f)	7	6	E	1
4(a)	19	2	E	2
(b)	19	2	E	3
(c)	19	2	E	3
(d)	15	2	D	3
(e)	15	2	E	3

Item	Topic	Theme	Activity	Ability
5(a)	2	3	C	3
(b)	19	2	C	2
(c)	19	2	C	2
(d)	13	1	B	2
(e)	5/19	1	A	3
6(a)	5	2	E	2
(b)	5	1	B	2
(c)	5	1	B	3
(d)	5	1	B	4
(e)	9/13	1	B	2
(f)	5	1	B	2
7(a)	3	1	B	4
(b)	3	1	B	4
(c)	3	1	B	4
(d)	3	2	D	2
(e)	3	2	D	2
8(a)	9/13	3	A	1
(b)	9/13	1	A	3
(c)	9/13	2	A	3
(d)	9/13	1	B	2
(e)	9/13	1	A	1
(f)	9/13	2	A	4

APPENDIX 4.

SUMMER 1972, NUFFIELD 'A' LEVEL G.C.E. CHEMISTRY PAPER 2. EXAMINER ____________________

CENTRE NUMBER	CANDIDATE NUMBER	QUESTION 1	QUESTION 2	QUESTION 3	QUESTION 4	QUESTION 5	QUESTION 6	QUESTION 7	QUESTION 8	QUESTION SUB-TOTALS								Paper Total
		a b c d e f	a b c d e f	a b c d e f	a b c d e	b c d e	a b c d e f	a b c d e	b c d e f	1	2	3	4	5	6	7	8	

1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59 61 63 65 67 69 71 73 75

APPENDIX 4 (Contd.) A. LEVEL G.C.E. NUFFIELD CHEMISTRY PAPER 3B

EXAMINER A NAME	INDEX NO.	EXAMINER B NAME	INDEX NO.

CENTRE NUMBER	CANDIDATE NUMBER		PAPER TOTAL		QUESTION SCORES																	
					1		2		3		4		5		6		7		8		9	
			A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B	A	B

CARD COLUMNS 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34

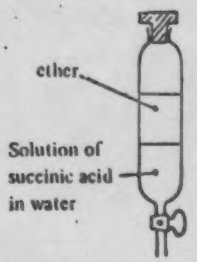
APPENDIX 5.

Do not write in this margin.

MARK | CARD COL.

1. Succinic acid is soluble in both water and ether, whereas water and ether are almost insoluble in each other. An experiment was devised to investigate an equilibrium between a solution of succinic acid in water and a solution of succinic acid in ether when the two solutions were in contact.

The apparatus used was a stoppered separating funnel and titration apparatus. A known volume of a standard solution of succinic acid in water was put into the funnel and ether added, as shown in the diagram:



1 mark (*a*) What would you do to make sure that equilibrium was established as soon as possible?

13

3 marks (*b*) After making sure that an equilibrium is established, a sample of the *aqueous* layer is to be separated for analysis. What precautions would you take, in the extraction of the sample, to avoid inclusion of some of the ether layer with the aqueous sample?

14

2 marks (*c*) Would the removal of the sample disturb the equilibrium between the succinic acid in the water layer and that in the ether layer? Briefly explain your answer.

15

3 marks (*d*) The dissociation constant, K_a, for succinic acid is $6 \cdot 33 \times 10^{-5}$ mol dm^{-3}. The aqueous sample is to be titrated against a standard solution of sodium hydroxide. Which of the following indicators would be most suitable for this titration? Briefly explain the reasons for your choice.

Thymol blue	working pH range	8·0 to 9·5.
Para-nitrophenol	„ „ „	5·5 to 7·6.
Bromo-phenol blue	„ „ „	2·8 to 4·6.

16

APPENDIX 6.

SOME ANCILLARY DATA

KEY

P1 - The Objective Test

P2 - The Structured Paper

P3A - Special Studies

P3B - Free Response Paper

TAP - Teachers' Practical Score

MPA - Moderated Practical Score

TEA - Teachers' Estimate for Overall Performance

GT - Grand Total

UGT - Unbiased Grand Total

1. Means and Standard Deviations of Various Parts of the Examination.

	MAX	MEAN			S.D.			MEAN/MAX.			S.D./MAX.		
	1970/ 71/2	1970	1971	1972	1970	1971	1972	1970	1971	1972	1970	1971	1972
P1	50	25.0	25.4	27.6	7.0	6.5	7.6	.50	.51	.55	.14	.13	.15
P2	100	52.0	56.4	48.7	15.2	15.3	14.6	.52	.56	.49	.15	.15	.15
P3A	20	-	9.1	9.0	-	3.6	3.4	-	.45	.45	-	.18	.17
P3B	30	-	15.0	13.7	-	4.3	4.6	-	.50	.46	-	.14	.15
MPA	100	69.0	68.0	67.4	13.2	13.1	13.5	.69	.68	.67	.13	.13	.14
GT	100	51.9	54.6	52.6	12.3	11.7	12.4	.52	.55	.53	.12	.12	.12
TAP	100	65.6	64.7	65.1	9.8	10.8	10.2	.66	.65	.65	.098	.11	.10
TEA	100	53.9	53.1	-	12.5	12.9	-	.54	.53	-	.12	.13	-

N.B. Maximum refers to raw scores. The actual weighting of the various parts of the examination was:

P1 - 30%, P2 - 30%, P3(A+B) - 25%, MPA - 15%

2. Correlations of Parts of the Examination

N.B. The correlations of each part score with the grand total are biased because the grand total includes the part score. The correlation with UGT gives the correlation of the part scores with the score on the remainder of the examination.

1970

	P1	P2	P3	MPA	TAP	TEA	GT
P1	-						
P2	.76	-					
P3	.54	.65	-				
MPA	.51	.54	.43	-			
TAP	.49	.52	.36	.71	-		
TEA	.73	.75	.60	.57	.59	-	
GT	.87	.92	.81	.67	.59	.81	-
UGT	.73	.82	.64	.56	.52	-	-

1971.

	P1	P2	P3A	P3B	MPA	TAP	TEA	GT
P1	-							
P2	.76	-						
P3A	.39	.44	-					
P3B	.54	.59	.47	-				
MPA	.53	.62	.37	.48	-			
TAP	.45	.50	.29	.39	.65	-		
TEA	.67	.73	.36	.51	.62	.64	-	
GT	.88	.92	.56	.76	.73	.57	.76	-
UGT	.75	.80	.49	.64	.63	.52	-	-

1972.

	P1	P2	P3A	P3B	MPA	TAP	GT
P1	-						
P2	.82	-					
P3A	.41	.44	-				
P3B	.55	.56	.43	-			
MPA	.56	.62	.31	.46	-		
TAP	.46	.49	.26	.37	.65	-	
GT	.91	.92	.54	.74	.72	.56	-
UGT	.80	.83	.48	.60	.63	.51	-

3. Estimates of Reliability of Various Parts of the Nuffield 'A' Level Chemistry Examination.

α = Cronbach's alpha coefficient of internal consistency

r = Estimate from highest unbiased correlation with other parts of the examination.

T = Backhouse formula for estimating the reliability of tests in which candidates have a choice of questions.

Written Papers

Year	Paper	α	r
1970	1	0.80	0.76
1971	1	0.75	0.76
1972	1	0.83	0.82
1970	2	0.85	0.82
1971	2	0.88	0.80
1972	2	0.88	0.83
		T	r
1970	3	-	0.65
1971	3B	0.61	0.64
1972	3B	0.55	0.60

Practical Work

Year	Assessment	α	r
1970	Moderated Score	0.85	0.56
1971	"	0.86	0.63
1972	"	0.89	0.63
1970	Teachers' Score	0.78	0.52
1971	"	0.83	0.52
1972	"	0.84	0.51

An estimate of the reliability of the subject grand total was made using Mosier's formula as suggested by Backhouse in Nuttall and Willmott; British Examinations, 1972, p.93. The results were:

1970	0.92
1971	0.93
1972	0.93

Activity B (Moderating sub-test on practical work)

Year	α	N (No. of Paper 1 and Paper 2 items.)
1970	.56	20
1971	.50	20
1972	.55	20

APPENDIX 7. QUESTIONNAIRES

UNIVERSITY OF LANCASTER

Department of Educational Research,
Cartmel College,
Bailrigg,
LANCASTER.

Tel: Lancaster 65201.

JCM/JS.

April, 1972.

For the attention of the Head of the Chemistry Department

Dear Colleague,

Nuffield 'A' Level Chemistry Examination

The Nuffield examination has many novel features, particularly the internal assessment of practical work by teachers. The examination was designed after close consultation with teachers during the school trials of the new course. But that was nearly five years ago and we are now investigating how well the examination is fulfilling its intended function. Closer co-operation between teachers and examiners has been one of our principal aims, and I would like to continue this by asking for your views on some aspects of our scheme for the assessment of practical work.

To this end, Geoffrey Sleightholme is conducting a survey of all schools who presented candidates for the 1971 examination, and copies of his questionnaire are attached.

I am sure you will agree that consultation of this sort is important, particularly when new examination techniques are introduced. I hope to get a substantial response, so I should be very grateful for your co-operation. Could you please distribute the questionnaires to all teachers who are teaching or have taught Nuffield 'A' Level Chemistry. It would be particularly helpful if you could let us have the completed questionnaires (stamped addressed envelope enclosed) returned by 8th May. All replies from individual schools will be regarded as confidential. We are also sampling student opinion, but your school did not fall within the random sample of students.

Yours sincerely,

J. C. Mathews

J.C. Mathews
Project Director.

UNIVERSITY OF LANCASTER

Department of Educational Research,
Cartmel College,
Bailrigg,
LANCASTER.

JCM/JS.

Tel: Lancaster 65201.

April 1972.

'A' Level Nuffield Chemistry

This questionnaire has been sent to teachers at ALL schools who submitted candidates for the Nuffield Advanced Level Chemistry Examination in 1971.

Its purpose is to obtain more definite knowledge about the present form of assessment of pupils' practical work and to discover in detail what problems (if any) exist. A different questionnaire will be going out to a representative sample of pupils. Your pupils may be amongst them.

It is fully realised that time is valuable to all teachers receiving this questionnaire, and for this reason questions have been kept as short and concise as possible. Obviously additional comments will be most welcome where this is felt appropriate, and a few of the questions do specifically ask for them. Space for more general comments is available on page 6. . Your help in this way will be much appreciated and your individual answers will be treated in absolute confidence.

Could you please return the completed questionnaire no later than 8th May, addressed to:

G. Sleightholme Esq., B.Sc., F.R.I.C.
Department of Educational Research,
Cartmel College,
University of Lancaster.
Bailrigg. LANCASTER.

/ A stamped addressed envelope is enclosed for this purpose.

PRELIMINARY NOTES:

Teachers may like to be reminded that the current system for practical assessment was established from the results of a questionnaire to teachers in 1967. The aims of the present enquiry can be realised by indicating that:-

Questions 1 - 8 are concerned with the principles and operation of the present system.

/ (Here you may wish to refer to the 'Notes on assessment of practical work by teachers' and a copy of these notes is enclosed for your convenience).

Questions 9 - 13 are concerned with the question of internal assessment and student/teacher relations.

Questions 14 - 21 are concerned with changing the system.

To record your responses to each question, please put a CIRCLE round the appropriate letter in the RESPONSE MARGIN. Do not hesitate to use pen, biro, pencil or whatever you prefer to write with for this purpose.

TEACHER QUESTIONNAIRE

RESPONSE MARGIN

1. The currently accepted objectives for the Nuffield Advanced Level Chemistry practical and their relative weightings for assessment purposes are as follows:

   - i Skill in observation (25%)
   - ii Ability to interpret observations (15%)
   - iii Ability to plan experiments (10%)
   - iv Manipulative skills (30%)
   - v Attitudes to practical work (20%)

   Assuming that these five objectives continue to be accepted for assessment purposes, do you consider that the weighting attached to:

   1.1. Skill in observation -

   A. Is much too high

   B. Is too high

   C. Is about right — A B C D E

   D. Is too low.

   E. Is much too low.

   1.2. Ability to interpret observations — A B C D E

   1.3. Ability to plan experiments — A B C D E

   1.4. Manipulative skills — A B C D E

   1.5. Attitudes to practical work — A B C D E

2. In your experience is a MEANINGFUL assessment of -

   2.1. Skill in observation -

   X. Straightforward Y. difficult. Z very difficul — X Y Z

   2.2. Ability to interpret observations — X Y Z

   2.3. Ability to plan experiments — X Y Z

   2.4. Manipulative skills — X Y Z

   2.5. Attitudes to practical work — X Y Z

3. How many students do you normally arrange to assess in the same practical session?

   X One of them Y. Some of them. Z. All of them. — X Y Z

4. Do your students work in pairs (or larger groups) when carrying out experiments which have been chosen for assessment?

   X. Never Y. Infrequently Z. Frequently — X Y Z

RESPONSE MARGIN

Do you consider the task of completing pupils' record cards -

X No extra burden,

Y. An extra burden falling inside normal teaching duties

Z. An extra burden falling outside normal teaching duties.

X Y Z

Please state any other effects which this form of assessment has on your normal work load.

In assessing a given experiment, do you find any difficulty in disregarding your previous knowledge of the pupil's practical ability?

X No difficulty or almost no difficulty

Y Some difficulty

Z Serious difficulty

X Y Z

Do you agree with the assertion that the present scheme for teacher assessment helps to sharpen the aims of chemistry practical work in general?

A. Strongly agree

B. Agree

C. Not sure

D. Disagree

E. Strongly disagree.

A B C D E

Do you consider that the Examining Board's instructions for the assessment of practical work give you sufficient freedom to assess as you think appropriate?

YES / NO

If NO please comment.

RESPONSE MARGIN

9. To what extent does oral communication between students and yourself play a part in your assessment?

X a substantial part
Y a minor part
Z no part.

X Y Z

10. Do you inform your students early in the course about the following?

(i) objectives of the practical assessment — YES / NO

(ii) proportion of marks allocated to each objective. — YES / NO

(iii) contribution of practical assessment to the total score. — YES / NO

(iv) the form of moderation of the teachers' marks. — YES / NO

11. Do you normally tell your pupils of your intention to assess a particular experiment? — YES / NO

If YES please answer 11. 1 and 11. 2.

11. 1 To what extent have you found student/teacher relations affected when students are carrying out an experiment which has been selected for assessment?

A distinctly improved
B. somewhat improved
C. not affected.
D. somewhat impaired
E. distinctly impaired.

A B C D E

11.2. To what extent do you feel that there is a general difference in performance in practical work when pupils are working on an experiment selected for assessment?

A. a marked improvement.
B. some improvement
C. no difference
D. some deterioration
E. a marked deterioration.

A B C D E

RESPONSE MARGIN

To what extent do you feel that student/teacher relations are affected in general by the fact that you are one of their examiners?

A. distinctly improved
B. somewhat improved.
C. not affected.
D. somewhat impaired.
E. distinctly impaired.

A B C D E

Do you feel that it is difficult to ignore general opinions about students (i.e.opinions based on general observations both inside and outside the classroom) in arriving at an assessment of practical work?

X. no difficulty or almost no difficulty
Y. some difficulty
Z. serious difficulty

X Y Z

In the light of your experience would you prefer

W. some form of internal assessment of practical work?
X. no internal assessment of practical work, but a practical test at the end of the course.
Y. a practical test at the end of the course.
Z. no assessment of practical work.

W X Y Z

If you have answered W, is the present form of assessment of the Nuffield Advanced Level Practical Chemistry -

F. satisfactory as it stands
G. satisfactory but requiring minor alterations
H. satisfactory in principle, but requiring major alterations.

F G H

If you have answered G or H and have modifications to propose do you wish to delete any of the five objectives? YES / NO

If YES please specify:

RESPONSE MARGIN

14. contd. If you wish to add any major objectives for assessment purposes please specify:

15. Would it facilitate your assessment if the major objectives were more precisely defined? YES / NO

If YES please specify.

16. Do you propose any change in the specified areas of work (i.e. changes, equilibria, kinetics, energetics)? YES / NO

If YES please comment.

17. Do you propose any change in the prescribed number of eight experiments for assessment? YES / NO

If YES please comment.

RESPONSE MARGIN

What is your opinion of the suggestion that assessment by impression (based on present objectives) of the WHOLE Sixth Form chemistry practical would be preferable to the existing system?

A. strongly agree

B. agree

C. not sure

D. disagree

E. strongly disagree.

A B C D E

How far do you agree with the assertion that whenever assessment of experimental work is intended, the students should always specifically be told?

A. strongly agree

B. agree

C not sure

D. disagree

E. strongly disagree.

A B C D E

How far do you agree with the assertion that any form of internal assessment of practical work should, as a matter of principle, be moderated by external means?

A. strongly agree

B. agree

C not sure

D. disagree

E. strongly disagree.

A B C D E

Would you be prepared for internal assessments of practical work that you had made to be communicated to the students?

i.	after scaling and moderating?	YES:	NOT SURE:	NO:
ii.	without scaling and moderating?	YES:	NOT SURE:	NO:

Any further comments:-

Signed...

School/College

UNIVERSITY OF LANCASTER

Department of Educational Research,
Cartmel College,
Bailrigg,
LANCASTER.

Tel: Lancaster 65201.

JCM/JS.

April, 1972.

Nuffield 'A' Level Chemistry

You will probably know that your 'A' level examination mark for practical chemistry is largely based on your teacher's assessment of your performance. This approach to examining is still rather new and the object of this enquiry is to get more definite knowledge of how it works in practice.

All teachers in the Scheme have received another questionnaire designed for the same purpose, and you are one of a chosen sample of students whose opinion we seek in this connection.

To record your responses to each statement, please put a circle round the appropriate letter in the RESPONSE MARGIN. Choose the response which is closest to your own opinion and the way this form of assessment affects you personally. All information from individuals will be kept confidential.

QUESTIONNAIRE TO STUDENTS

RESPONSE MARGIN

1. This form of assessment is preferable to a single practical examination at the end of the course.

A. Strongly agree B. Agree C. Uncertain
D. Disagree E. Strongly disagree

A B C D E

2. This form of assessment encourages more care than might otherwise be the case with practical work.

A. Strongly agree B. Agree C. Uncertain
D. Disagree. E. Strongly disagree.

A B C D E

3. This form of assessment is a source of anxiety in school work.

A. Strongly agree. B. Agree. C. Uncertain
D. Disagree E. Strongly disagree.

A B C D E

4. This form of assessment has a tendency to change student/teacher relations. - for the better:

A. Strongly agree. B. Agree C. Uncertain.
D. Disagree E. Strongly disagree.

A B C D E

5. This form of assessment has a tendency to change student/teacher relations - for the worse:

A. Strongly Agree. B. Agree C. Uncertain
D. Disagree E. Strongly disagree

A B C D E

6. This form of assessment tends to make one feel self-conscious in doing practical work.

A. Strongly agree B. Agree. C. Uncertain.
D. Disagree. E. Strongly disagree.

A B C D E

7. This form of assessment helps to make clear the aims of each experiment.

A. Strongly agree. B. Agree. C. Uncertain.
D. Disagree. E. Strongly disagree.

A B C D E

8. Occasions of assessment in practical work reduce fruitful co-operation between students.

A. Strongly agree. B. Agree. C. Uncertain.
D. Disagree. E. Strongly disagree.

A B C D E

Continued Overleaf....

RESPONSE MARGIN

9. Students should always be informed when they are being assessed on a particular occasion.

A. Strongly agree. B. Agree. C. Uncertain.
D. Disagree E. Strongly disagree.

A B C D E

10. Any form of assessment attempting to measure practical work is bound to be rather inaccurate.

A. Strongly agree. B. Agree. C. Uncertain
D. Disagree E. Strongly disagree.

A B C D E

11. Briefly give any reasons, not expressed in the above answers, that you see:

(a) in favour of the present system of assessment -

(b) against the present system of assessment.

TEACHERS OF 'A' LEVEL NUFFIELD CHEMISTRY

Department of Educational Research,
Cartmel College,
The University,
LANCASTER.
April 1973

Dear Colleague,

GCE 'A' level Nuffield Chemistry Examination Project

The design and function of this examination have relied heavily on consultation and co-operation with teachers. Our research project seeks to continue this joint activity between schools and examiners and last year we were very pleased at the high response to our questionnaire on the Assessment of Practical Work, the results of which have aroused much interest.

We now turn to some aspects of the course and the written examination in the attached questionnaire. This is for all teachers who have prepared students for the GCE 'A' level Nuffield Chemistry examination and we hope sufficient copies are enclosed for all such teachers in your school. Please write to us if more are needed.

In order not to take up to much of your time, we have kept this questionnaire as concise as possible. Nonetheless we hope that you will make further comments if you so wish, and have left the reverse side of each page clear for that purpose.

Your help in this investigation is greatly appreciated and your answers will be treated with the utmost confidentiality. A stamped addressed envelope is provided for each questionnaire; we hope you will be able to return the questionnaire within two weeks. We shall send you a copy of our report on this survey as soon as possible.

In addition, we think it important to sound STUDENT OPINION on some aspects of the course and enclose a shorter questionnaire for Students taking the examination this summer. This is only being sent to a sample of schools, so we hope all your Students will complete it.

Yours sincerely,

J C. Mathew

J C Mathews
Project Director

NOTES ON COMPLETING THE ATTACHED TEACHER QUESTIONNAIRE:

1. With the exception of question C2, we are concerned with 'A' level NOT 'O' level.
2. We frequently refer to the "published Guides", meaning the series of books published by Penguin Books for the Nuffield Foundation as Guides to Teachers and to Students on Nuffield Advanced Chemistry.
3. Please answer questions A2 and C1 by writing in the small box provided.
4. Most questions have a set of responses; please choose only ONE reponse.
5. Please indicate your response by drawing a small ring around the appropriate response code number (see example). This should be the response best resembling your own views on the question. Please feel free to make further comments on the reverse side of the page.

MPLE: 9. What is your present occupation?

School Teacher .... [✓ 1]
Banker 2
Dentist 3

SECTION A

	COMPUTER CARD COLUMNS
☐☐☐☐☐☐	1 - 6

Details of your school:

(a) TYPE OF SCHOOL	Maintained Grammar	1	7
	Technical Grammar	2	
	Comprehensive	3	
	Direct Grant	4	
	Independent	5	
	Other (please specify on the left)	6	
(b) SEX(ES) PRESENT	Boys only	1	8
	Mixed	2	
	Girls only	3	
(c) NUMBER OF STUDENTS	Up to 300	1	9
	301 - 500	2	
	501 - 600	3	
	601 - 800	4	
	Over 801	5	

How many years have you been teaching ADVANCED level NUFFIELD Chemistry (to the nearest whole number)? .... ☐ 10

From the point of view of its MERITS AS A 6th FORM CHEMISTRY CURRICULUM, how would you describe the published 'A' level Nuffield Course:

Should be completely rewritten	1	11
Needs major changes	2	
Satisfactory after minor changes	3	
Satisfactory in its present form	4	
Excellent in its present form	5	

In teaching NUFFIELD 'A' level Chemistry, do you find that the requirements of the NUFFIELD GCE 'A' level examination CONSTRAIN you when deciding:

(a) WHAT TO TEACH?	Yes, to a large extent	1	12
	Yes, to some extent	2	
	Yes, to a small extent	3	
	Not significantly	4	
(b) HOW TO TEACH IT?	Yes, to a large extent	6	13
	Yes, to some extent	7	
	Yes, to a small extent	8	
	Not significantly	9	

SECTION B

We'd like to ask two questions about each TOPIC of the Advanced Nuffield Course:

1. How closely do you currently follow the published Guides in teaching each TOPIC?
2. Referring to the TOPICS as described in the published Guides, how difficult do you think your students find each TOPIC (relative to the average level of difficulty of the Course)?

If, in some TOPICS, you depart so far from the Guides that you can't answer question 2, then omit question 2 for those TOPICS.

Otherwise, please ring ONE code number for each question, for each TOPIC.

	QUESTION ONE					QUESTION TWO					COMPUTER CARD COLUMNS	
TOPIC	Without significant change	With minor modifications	With major modifications	Guides not followed	Topic not taught	Much harder than average	Harder than average	ABOUT AVERAGE	Easier than average	Much easier than average		
AMOUNT OF SUBSTANCE	1	2	3	4	5	1	2	3	4	5	18	19
PERIODICITY	1	2	3	4	5	1	2	3	4	5	20	21
THE MASSES OF MOLECULES & ATOMS; THE AVOGADRO CONSTANT	1	2	3	4	5	1	2	3	4	5	22	23
ATOMIC STRUCTURE	1	2	3	4	5	1	2	3	4	5	24	25
THE HALOGENS & OXIDATION NUMBERS	1	2	3	4	5	1	2	3	4	5	26	27
THE s-BLOCK ELEMENTS; & THE ACID-BASE CONCEPT	1	2	3	4	5	1	2	3	4	5	28	29
ENERGY CHANGES & BONDING	1	2	3	4	5	1	2	3	4	5	30	31
STRUCTURE & BONDING	1	2	3	4	5	1	2	3	4	5	32	33
CARBON CHEMISTRY, part 1	1	2	3	4	5	1	2	3	4	5	34	35
INTERMOLECULAR FORCES	1	2	3	4	5	1	2	3	4	5	36	37
SOLVATION	1	2	3	4	5	1	2	3	4	5	38	39
EQUILIBRIA: GASEOUS & IONIC	1	2	3	4	5	1	2	3	4	5	40	41
CARBON CHEMISTRY, part 2	1	2	3	4	5	1	2	3	4	5	42	43
REACTION RATES	1	2	3	4	5	1	2	3	4	5	44	45
EQUILIBRIA: REDOX & ACID-BASE SYSTEMS	1	2	3	4	5	1	2	3	4	5	46	47
SOME d-BLOCK ELEMENTS	1	2	3	4	5	1	2	3	4	5	48	49
EQUILIBRIUM & FREE ENERGY	1	2	3	4	5	1	2	3	4	5	50	51
CARBON COMPOUNDS WITH LARGE MOLECULES	1	2	3	4	5	1	2	3	4	5	52	53
SOME p-BLOCK ELEMENTS	1	2	3	4	5	1	2	3	4	5	54	55

COMP
CARD
COLU

SECTION C

This final section is ONLY for teachers who now teach NUFFIELD 'A' level Chemistry but previously taught Chemistry to students taking any of the other GCE 'A' level Chemistry examinations. We shall call these other examinations and their syllabuses "TRADITIONAL".

1. How many years did you teach TRADITIONAL 'A' level Chemistry (to the nearest whole number; if more than 9 just put 9)? .... ☐ 56

2. Was NUFFIELD 'O' level Chemistry already being taught in your present school when NUFFIELD 'A' level Chemistry was adopted there? 57
   - Yes 1
   - No 2
   - Don't know .. 3

3. Please ring whichever ONE of the following statements best describes how you began teaching NUFFIELD 'A' level Chemistry: 58
   - Your school adopted NUFFIELD 'A' level at your suggestion 1
   - You supported a decision to adopt NUFFIELD 'A' level 2
   - You were neutral with respect to a decision to adopt NUFFIELD 'A' level ... 3
   - You disagreed with a decision to adopt NUFFIELD 'A' level 4
   - You moved to a school & you were pleased that it taught NUFFIELD 'A' level . 5
   - You moved to a school which happened to teach NUFFIELD 'A' level 6
   - Some other statement, please specify overleaf 7

In the following questions, please compare your experience of the NUFFIELD 'A' level Chemistry written examinations with your experience of TRADITIONAL 'A' level Chemistry written examinations. (We are not concerned here with practical exams.)

We shall ask if you think any difference exists between them in various respects.

If you think none exists, circle the "no difference" code number.
"Slightly" implies a small but definite difference.
"Markedly" implies a large difference

Please ring ONE code number for each question.

4. On the whole, WHICH WRITTEN EXAMINATION SYSTEM:	NUFFIELD (markedly)	NUFFIELD (slightly)	NO DIFFERENCE	TRADITIONAL (slightly)	TRADITIONAL (markedly)	
(a) laid the greater stress on REMEMBERING FACTS	1	2	3	4	5	59
(b) laid the greater stress on UNDERSTANDING of CONCEPTS	1	2	3	4	5	60
(c) laid the greater stress on APPLYING experience in solving unfamiliar problems	1	2	3	4	5	61
(d) tested the more COMPREHENSIVE selection of your students' work	1	2	3	4	5	62
(e) produced the greater PRE-EXAMINATION TENSION amongst students.	1	2	3	4	5	63
had the greater CONSTRAINING effect on your teaching	1	2	3	4	5	64
was easier for your students to pass	1	2	3	4	5	65

ADVANCED LEVEL NUFFIELD CHEMISTRY STUDENTS

Department of Educational Research,
Cartmel College,
UNIVERSITY OF LANCASTER.
April 1973

To Students taking the 'A' level
Nuffield Chemistry Examination in
Summer 1973

The 'A' level Nuffield Chemistry Course and examination were designed in close consultation with schools and now in the follow-up phase we seek to continue this School-Examiner co-operation and hope that you will help us by completing this questionnaire. We know that your time is precious now that the GCE exams are approaching, but we can only ask these questions towards the end of the course.

We have kept the questions as concise as possible. Overleaf we ask for your views on the DIFFICULTY and on the INTEREST of each Topic of the Course, and on this page for some background information.

Your help is greatly appreciated.

HOW TO COMPLETE THE QUESTIONNAIRE:

1. If a question has a set of alternative responses, choose only ONE response.
2. Indicate your response by drawing a small ring around the appropriate code number (see example).
3. The response you indicate should be that most closely resembling your own opinion on the question. We would welcome any further comments you may wish to make on a separate sheet of paper.

PLE: 9. Are you studying Advanced level Nuffield Chemistry? Yes ... [✓ 1]
No 2

		OFFICE USE ONLY
	☐☐☐☐	1 - 4
Are you studying Mathematics at 'A' level? (Any "maths" subject counts e.g. "PURE", "FURTHER")	Yes .... 1 No 2	5
How many GCE 'O' level passes did you have upon entering the 6th form .. (please write the number in the box opposite)	☐	6 - 7
What sex are you?	FEMALE ...... 1 MALE 2	8
Apart from Chemistry, do you intend to take any other NUFFIELD 'A' level examinations? yes, please list the subject(s)	Yes .... 6 No 7	9

STUDENT QUESTIONNAIRE - SIDE 2

We'd like to ask 2 questions about each TOPIC of the Advanced Nuffield Course:

1. In relation to your overall interest in Chemistry, how interesting do you find each TOPIC of the 'A' level NUFFIELD Chemistry Course?
2. In relation to your understanding of the overall difficulty of the 'A' level NUFFIELD Chemistry Course, how difficult do you find each TOPIC?

Please give your answers by ringing ONE code number for each question for each TOPIC

If you can't answer for any particular TOPIC (perhaps because you haven't yet studied it), don't ring any code number at all for that TOPIC.

	TOPIC	QUESTION ONE					QUESTION TWO					OFFICE USE ONLY
		MUCH LESS interesting	LESS interesting	OF AVERAGE interest	MORE interesting	MUCH MORE interesting	MUCH EASIER than average	EASIER than average	AVERAGE	HARDER than average	MUCH HARDER than average	
1	AMOUNT OF SUBSTANCE	1	2	3	4	5	1	2	3	4	5	10 11
2	PERIODICITY	1	2	3	4	5	1	2	3	4	5	12 13
3	THE MASSES OF MOLECULES & ATOMS; THE AVOGADRO CONSTANT	1	2	3	4	5	1	2	3	4	5	14 15
4	ATOMIC STRUCTURE	1	2	3	4	5	1	2	3	4	5	16 17
5	THE HALOGENS & OXIDATION NUMBERS	1	2	3	4	5	1	2	3	4	5	18 19
6	THE s-BLOCK ELEMENTS; & THE ACID-BASE CONCEPT	1	2	3	4	5	1	2	3	4	5	20 21
7	ENERGY CHANGES & BONDING	1	2	3	4	5	1	2	3	4	5	22 23
8	STRUCTURE & BONDING	1	2	3	4	5	1	2	3	4	5	24 25
9	CARBON CHEMISTRY, part 1	1	2	3	4	5	1	2	3	4	5	26 27
10	INTERMOLECULAR FORCES	1	2	3	4	5	1	2	3	4	5	28 29
11	SOLVATION	1	2	3	4	5	1	2	3	4	5	30 31
12	EQUILIBRIA: GASEOUS & IONIC	1	2	3	4	5	1	2	3	4	5	32 33
13	CARBON CHEMISTRY, part 2	1	2	3	4	5	1	2	3	4	5	34 35
14	REACTION RATES	1	2	3	4	5	1	2	3	4	5	36 37
15	EQUILIBRIA: REDOX & ACID-BASE SYSTEMS	1	2	3	4	5	1	2	3	4	5	38 39
16	SOME d-BLOCK ELEMENTS	1	2	3	4	5	1	2	3	4	5	40 41
17	EQUILIBRIUM & FREE ENERGY	1	2	3	4	5	1	2	3	4	5	42 43
18	CARBON COMPOUNDS WITH LARGE MOLECULES	1	2	3	4	5	1	2	3	4	5	44 45
19	SOME p-BLOCK ELEMENTS	1	2	3	4	5	1	2	3	4	5	46

APPENDIX 8.

APPENDIX 8

The measurement of objectives in conjunction with certain experiments